DOCUMENT RESUME

ED 238 903 TM 830 851

AUTHOR	Fillmore, Charles J.; Kay, Paul
TITLE	Text Semantic Analysis of Reading Comprehension Tests. Final Report.
INSTITUTION	California Univ., Berkeley. Inst. of Human Learning.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	31 Aug 83
GRANT	NIE-G-790121
NOTE	79p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Educational Technology; Elementary Secondary Education; Evaluation Methods; Interviews; Language; *Reading Comprehension; Reading Difficulties; *Reading Processes; Reading Research; *Reading Tests; Research Problems; *Semantics; *Standardized Tests
IDENTIFIERS	*Berkeley Reading Tests Project; *Textual Analysis

ABSTRACT

The Berkeley Reading Tests Project has selected as its principal object of study standardized tests of reading comprehension. The interest is in the assessment of such tests, with respect to their performance as measures of the reading comprehension abilities of school children. A serious study of the assessment of reading comprehension requires an understanding of the process of reading comprehension itself, and such an understanding requires in turn an understanding of general language interpreting abilities. This study is aimed at (1) the practical goal of evaluating and improving the existing technology for assessing reading comprehension in American schools; (2) the goal of increasing the knowledge of the reading process itself, and thus perhaps contributing to the improvement of the reading curriculum; and (3) the scientific goal of understanding language as a peculiarly human ability. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED238903

Final Report to NIE: Text Semantic Analysis of Reading Comprehension Tests

Charles J. Fillmore
Paul Kay

with the active collaboration of:
Linda Coleman

Thomas W. Larsen
Catherine O'Connor

Staff:
Karen Carroll
Linda Coleman
Katherine Kovacic
Judith Langer
Thomas W. Larsen
Catherine O'Connor

August 31, 1983

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM CS

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Institute of Human Learning
University of California
Berkeley, California 94720

NIE Grant No. G-790121 Rev. 1

TM 830 857

Outline

1. Introduction

1.1 Tests as Texts and as Tasks

1.1.1 Instruments for Assessing Reading Comprehension

1.1.2 Reading Comprehension in General

1.1.3 Language Comprehension in General

1.2 A sample test: "Nasrudin"

1.3 The Purpose and Nature of our Interview

1.3.1 A Model interview

1.3.2 The Problem of Genre (in reading generally and in reading tests in particular)

2. The Ideal Reader Analysis and its Use in the Interviews

2.1 Column One: Notation of the IR's Processing of the Text

2.2 Column Two: Properties of the Interpreter

2.2.1 The Parsimony Principle

2.2.2 Additional Properties of the Interpreter

2.3 Column Three: Interview Questions

3. Results

3.1 Critique of Text Materials

3.1.1 Critique of Questions

3.1.2 Critique of Texts

3.1.3 The Pragmatics of the Reading Test Situation

3.2 Findings of the Interviews

APPENDICES

A. Passages Used in the Study

B. Dissemination

1. Introduction

The Berkeley Reading Tests Project has selected as its principal object of study standardized tests of reading comprehension. Our interest is in the assessment of such tests, with respect to their performance as measures of the reading comprehension abilities of the school children who are required to interact with them. We believe that any serious study of the assessment of reading comprehension requires an understanding of the process of reading compehension itself, and that such an understanding requires in turn, to an important though unknown degree, an understanding of general language interpreting abilities. Our study is thus aimed at (1) the practical goal of evaluating and improving the existing technology for assessing reading comprehension in American schools, (2) the more 'middle-range' goal of increasing our knowledge of the reading process itself, and thus perhaps contributing to the improvement of the reading curriculum, and (3) the more general scientific goal of understanding language processing as a peculiarly human ability.

1.1 Tests as Texts and as Tasks

1.1.1 Instruments for assessing reading comprehension

Research on the evaluation of reading comprehension has not for the most part been conducted within the framework of a theory of reading comprehension, nor even, as far as we have been able to tell, has such work been conducted in connection with attempts to develop such a theory. Decisions on the adequacy of reading test items are normally justified in terms of correlations across various groups of test subjects with regard to which answers to which items are selected by the same subjects. Often, demographic characteristics of the subjects are made part of the correlational matrix. The process of constructing reading test items, on the other hand, is treated as a private matter, inaccessible to study by the scholarly community. Whatever valuable lore may be possessed by the constructors of reading test items is not generally made available to the public. One cannot find documents accompanying particular reading tests which explain why one particular set of distractors has been preferred to another set for some particular question, why one formulation of a question is better than another for a given passage, on what explicit basis a certain question may be said to test 'inference' versus 'main idea' versus 'literal comprehension', etc.; but above all, there is no discussion of the suitability of the items as texts--that is, there is no discussion of the precise processing demands individual test items place on the readers who are expected to deal with them.

The relation between a text and the questions that are alleged to test comprehension of it remains entirely intuitive and unexplicated. If a text telling the story of Goldilocks were followed by a set of questions regarding the geography of South America, such an item might well pass all the statistical tests to

which reading test items are canonically subjected: correct answers on this item might correlate positively with correct answers on other items and might show no sex, race or regional bias, etc. Such a proposed item would, of course, be rejected on obvious, intuitive grounds, but it is precisely this kind of intuition that remains forever unexplicated under the existing technology of test construction and evaluation.

Our approach to the study of the assessment of reading comprehension has been to place this study within the framework of a study of the comprehension process itself--stressing those aspects of the comprehension process whose deficiencies in young readers concern us most.[1] Limiting ourselves to reading test items consisting of complete passages followed by questions, we take as our primary data the performance of individual readers on individual reading test items. The first problem we pose--regardless of how the reader responded to the questions--is through what steps and with what success the reader comprehended the passage presented in the item. An ability to evaluate the reader's comprehension performance presupposes a standard, correct, or 'ideal' interpretation of the passage, something with which the young reader's interpretation is to be compared. An analogy may be drawn to the assessment of arithmetical competence in children. The first step in assessing a child's competence in the solution of a set of, say, subtraction problems is to solve these problems ourselves! After this initial step is taken, the child's answers may be compared to the correct answers in a variety of ways to discover systematic gaps in the child's abilities to deal with this kind of problem, as illustrated in the work of Brown and Burton (1978). In order to evaluate a child's comprehension of a text it is necessary first to produce a standard or ideal interpretation of that text against which the child's interpretation can be judged. We see the omission of this essential first step in current research on reading comprehension and its evaluation as a serious flaw.

The interpretations of texts that we take as our standard are produced by a process of analysis that posits for each text an IDEAL READER of that text. The ideal reader is a hypothetical text processor optimally constructed to benefit from exposure to this particular text. The ideal reader knows everything that the text presupposes and is possessed of just that background knowledge and those processing abilities needed to extract from the text everything that it poses.

The ideal reader is the foundation of the implicit theory of

---

[1] We are not primarily concerned with young readers' problems in decoding and so the important issue of decoding is not stressed in our account of normal comprehension, though it must of course be noted.

reading comprehension within which we conduct our study of reading comprehension assessment. This is a strongly text-based theory of reading. (In de Beaugrande's (1981) classification, it exhibits a high degree of 'Utilization'.) Each text requires of its reader certain knowledge and abilities, and in general no two texts will determine the same ideal reader. Thus the ideal readers of texts (1) and (2) below are required to make use of distinct sets of background knowledge and are required to draw different kinds of inferences in making coherent the reported events and the consequent emotional states of the participants.

(1) Georgie asked Mom for two cookies. He was happy when she gave him three.

(2) Georgie asked Mom for two cookies. He was happy when she gave him one.

The ideal reader of text (2), but not text (1), is expected to know, or be able to imagine, that people sometimes ask for more than they really want in the hope of fully satisfying their wants via partial satisfaction of their demands. (Casual observation suggests that small children may employ this schema in the organization of their actual request behavior before they are able to employ it in the interpretation of texts about the request behavior of others.)

Implicit in our 'ideal reader' approach to reading comprehension and its assessment is the idea that reading is a constructive process. The ideal reader is equippped with a set of processes and an array of schemata. Confronting a text--reading--means bringing a subset of schemata to bear on the text, via an appropriate activation of processes, so as to produce an ENVISIONMENT of the text: the representation in the reader's mind of the world the text is taken to convey. To the extent that an actual reader's envisionment of a given text conforms to the envisionment of that text by an ideal reader, we say that adequate comprehension has occurred. To the extent that an actual reader's envisionment differs from that of the ideal reader, we have discovered a failure of comprehension.

Notice that defined in this way the ideal reader of a text is relativized to and largely determined by the text itself. It is thus possible that a given text will determine an ideal reader who is possessed of schemata or processing abilities which we think are either beyond the abilities of any real readers, beyond the abilities of the readers at whom the particular text is aimed, undesirable, or logically or psychologically impossible. For example, if a text requires clairvoyance of its ideal reader, or if it requires its reader to believe a contradiction, we consider it a flawed text. Similarly, texts that are known to be aimed at particular populations, e.g., school children, and which require of their ideal reader schemata or processing abilities that it is either unrealistic to impute to the target population or

undesirable that members of that population have, may also be judged as flawed. Unfortunately, almost all of the texts that we have encountered as actual tests of reading comprehension are flawed in one or more of these ways.[2]

In the main body of this report we shall have much more to say about the concept 'ideal reader' and the role it plays in the theoretical and methodological aspects of our study of reading tests. For the moment it suffices to remember that the concept of the ideal reader for a particular text furnishes a standard against which real readers of that text may be judged and also furnishes a standard against which the text itself may be judged: if the ideal reader it ordains is logically or psychologically impossible, or ill-matched to the intended readership of the text, or venal, or otherwise undesirable, the text is inadequate. If we are to be able to evaluate the extent to which existing tests of reading comprehension in fact test reading comprehension, then we must have a method of assessing reading comprehension independently of the scores on these tests. The concept of the ideal reader, it will be argued below, provides the beginning for the development of such a method.

In subsequent sections of this report, we will discuss aspects of the ideal reader which show it to be a device of considerable sophistication. There is, however, one important respect in which the ideal reader we posit is less sophisticated than able, mature readers. Our impoverished ideal reader is one that lacks the capacity to reflect upon the history of its own processing of a text. This TRUSTING IDEAL READER is literal minded. If it forms a hypothesis and then encounters evidence that leads it to reject that hypothesis, it doesn't have to ask itself second order questions such as, "Why would the author have put in the information that led me to the first hypothesis if he or she were only to give me conflicting information later? Is there some kind of literary trick going on here?" The trusting

---

[2] The testing industry makes a serious effort to keep these texts free of ethnic and sexual stereotypes, and none of the texts we have examined is defective by virtue of requiring the ideal reader to employ derogatory ethnic or sexual stereotypes in the construction of its envisionment. So far as we can tell, this is the only domain of content with respect to which the testing industry has successfully concerned itself. As an example of the kinds of flaws we have found, consider our discussion below of a test item concerned with the description of the first phonograph. Not one of over thirty third graders we interviewed recognized the word hi-fi and only one recognized the word phonograph, both of which occurred critically in the text, though all of them knew the words stereo and record player neither of which appeared in the text. Section 1.2 below presents further examples of textual flaws.

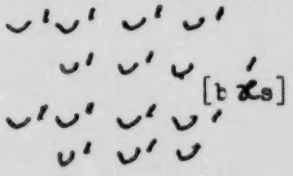
ideal reader is artless in the sense that it does not produce secondary interpretations of a text by reflecting upon the processing it has done in producing the primary interpretation. The trusting ideal reader does not construct an image of the author of the text it is processing and does not reflect upon this author's possible motivations.

The concept of the trusting ideal reader (or hearer) is readily illustrated by a very short text intended primarily to be heard rather than read, a bit of doggerel enjoyed by teenagers of certain times and places.

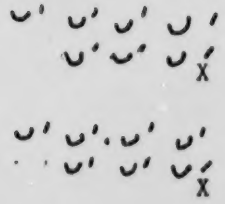
> Of all the things I'd rather be,
> I'd rather be a bass.
> I'd climb up to the tops of trees
> And slide down on my hands and knees.

Part of the intended humorous effect of the text depends on the whimsical image of fish climbing trees; we will leave this matter for the moment and return to it later. For the rest, the addressee who gets the point and intended humor of the text incorporates the trusting addressee and reflects upon the results of this very trustingness in arriving at the intended rhetorical effect.

From the point of view of the trusting addressee the analysis of the text goes in part as follows. At the end of the third line, one has noted that the rhyme scheme is so far a b c (or perhaps a b a, depending on whether trees is heard as an imperfect rhyme for be), and with regard to meter that all three lines are perfectly iambic, having four, three, and four feet respectively. The pattern of rhyme and meter so far perceived is as follows:



One therefore contracts the expectation that the next (i.e., the fourth) line will (1) end in a stressed syllable concluding with [-æs] to rhyme with bass and (2) contain three iambic feet, completing the pattern given below which is the basic ballad form, an extremely familiar genre in our culture.[3]



At the end of the word my in the fourth line--with one syllable left to go--these hypotheses are still working perfectly. We lack only a single stressed syllable to complete the ballad form (4 3 4 3 iambs), and there is a monosyllabic word which rhymes with bass, which names a part of the body one can slide on--thereby making as much semantic sense as one may hope for in a text containing the whimsy of bass climbing trees--and which, above all, provides the satisfaction of completing the pattern of rhyme, meter, and meaning with a tabooed word. For these reasons the trusting ideal hearer expects that word at this point.

When line 4 continues with hands and knees the trusting addressee cancels his expectation without further thought and notes with simple satisfaction that the poem has been successfully concluded with a rhyme scheme a b c c and a matching meter of 4,3,4,4 iambs. The trusting ideal addressee does not reflect upon the fact that his expectations have been violated, he merely revises them. He therefore doesn't appreciate that a trick has been played on him: that he has been led to contract an expectation for a tabooed word on the basis of the rhythmic, metric, and semantic pattern, but that that pattern has been completed satisfactorily without the use of the taboo word. The sophisticated addressee, on the other hand, appreciates that a joke has been played on her in that she has been convicted in her own mind of dirty-mindedness for having supplied the tabood word before it was produced, only to find that the poem was successfully (and innocently) completed without it.

What we want to note here is that the sophisticated ideal reader incorporates the trusting ideal reader. The sophisticated

---

[3] One of the many familiar quatrains that might be cited is from Wilde's Ballad of Reading Gaol:

> For each man kills the thing he loves;
> Let this by all be heard.
> Some do it with a bitter look.
> Some with a flattering word.

Note the rhyme of heard and word.

ideal reader gets the joke by virtue of noticing how the trusting ideal reader has been led down the garden path. It seems a useful working hypothesis that many if not all texts that require something more than our trusting ideal reader to appreciate their full rhetorical effect require as one component of their sophisticated ideal reader a trusting ideal reader. In so far as this hypothesis is correct, any theory that aims to explain the processes used by sophisticated readers in interpreting literary texts that rely on complex rhetorical strategies will need to include a theory of a trusting ideal reader. We therefore have grounds to hope that what we learn about the reading process by positing a trusting ideal reader who is competent to read very simple texts may be of use in understanding the workings of sophisticated readers who are able to interpret more complex texts successfully.

1.1.2 Reading comprehension in general

The kind of representation of the process and product of language comprehension which we aim to provide most closely resembles the work of various natural language understanding projects within the field of artificial intelligence. Of special relevance are the works of Roger Schank, Bonnie Meyers, Walter Kintsch, Carl Frederiksen, David Rumelhart, and Robert de Beaugrande.

The AI approaches mentioned differ from each other considerably, but they appear to us to be more or less in agreement on the following goals. They assume a theory with (1) a knowledge base comprising something like semantic memory plus whatever information the current text or the current situation has made available; (2) a defined set of processing rules or strategies (parsing algorithms, lexical look-up, schema-activation procedures, procedures for deriving inferences, etc.); and (3) an intuitively refined sense of what could constitute an appropriate representation of the content of a text, each such representation corresponding to what we refer to as the envisionment of the text's world. Such an account of a text is taken to be a structure of information about the text and its interpretation upon which certain 'cognitive' operations (such as questioning) can be performed. Progress in such projects comes about when the researchers (1) present the system with a text it has never confronted before, see whether the knowledge base provides, in the proper form, that information upon which the processing strategies can transform the text into a correct envisionment, and (2) check the correctness of the envisionment by seeing what questions about it the system can answer correctly. Whenever such tests fail, the researcher asks about the sufficiency of the existing knowledge base or the performance of the existing processing rules and modifies these in whatever ways seem to help. To give a simple example of this kind of "progress"-- if we were to find a text whose interpretation depended crucially on the reader's knowing that the number of planets in the solar system was nine and the current knowledge base failed to contain that information, then the designer of the

system would attempt to enhance the knowledge base by adding to it just that fact and introducing whatever 'paths' to that fact needed to exist.

Despite the overall similarity between our approach and the generalized AI approach we have just characterized (or caricatured) there are some differences. A consequence of the text based character of the ideal reader approach is its capability to be used as an evaluative standard for both actual readers and for texts. The usual AI program is based on the assumption that the text is sensible and appropriate and attempts to make what sense it can of the text. The ideal reader approach does not start out with any assumptions regarding the coherence of the text and is thus able to discover that it is being asked to believe contradictions or to perform in other ways that reflect negatively on the text that determines its behavior. Similarly, the AI approach counts a success when a reader-simulating program appears to act in the same way a good reader would, but it provides no standard against which less than perfect readers may be judged.

At the level of simulating or analyzing the individual text, our approach can also be constrasted with the general AI approach in terms of what is assumed and what is calculated from these assumptions.

The contrast here is methodological rather than theoretical, but it is nevertheless important. The generalized AI approach and the Ideal Reader (IR) approach share the following global characterization: A Text, in a Context, is submitted to an interpretive system consisting of a Knowledge Base and a set of Procedures to produce an Interpretation. (Since treatment of the Context does not enter into the story differentiating the two approaches, we do not mention it further here.) In practice the AI approach works as follows: a Knowledge Base and a set of Procedures are taken as given. These are presented with a Text. The system derives an interpretation in the sense of augmenting its initial Knowledge Base K to K'. This interpretation is then checked against the intuitively correct interpretation (e.g., by asking the system containing the new Knowledge Base K' questions that we think a competent human could answer correctly only after reading the Text.) In so far as the Interpretation is found lacking, adjustments are made to the Knowledge Base, the Procedures, or both. Another cycle is begun with the same or a different Text.

The IR approach proceeds differently. Only P, but not K, is taken as given. P is then presented with T and I. What the system 'derives' is Kt, the Knowledge Base required of the ideal reader of this particular Text. The system also keeps track of the proper subset Pt of P that consists in the set of procedures used in the processing of this particular Text. Together Kt and Pt comprise the Ideal Reader of the text T. (The Ideal Reader of T, representable as

$$IRt = \{Kt, Pt\} \quad (Kt \subset K, \quad Pt \subset P)$$

is then usable as a standard of evaluation both for T itself and for any real reader of T.) The differences of approach are summarized in the following table.

AI	IR
Given: P, K	Given: P
Presented: T	Presented: T, I
Derive: I (=K'-K)	Derive: Kt
	Note: Pt
Evaluate: I, by, e.g., questioning K'.	Evaluate: T by comparing real reader with [Kt, Pt]

One further difference between the IR approach and most, if not all, AI approaches, is that the former operates on the text in a linear fashion in calculating the interpretation. That is, the text is taken chunk by chunk, and the system seeks to simulate what the Ideal Reader has in mind at each point in reading the text. The Ideal Reader is thus equipped to register expectation, surprise, confirmation, suspicion and other temporally oriented cognitive states. A final distinction in practice is that the IR approach proceeds manually, not by automatic computation, and is therefore able to employ procedures that lack formal specification; this feature has both obvious advantages and obvious disadvantages.

Thus our text-based approach to the reading process meshes with our need to study actual texts and actual readers to produce the methodology based on the ideal reader concept introduced above and discussed in greater detail below. The ideal reader sets a standard against which both the real reader and the text may be evaluated, and our research to date has confirmed our initial suspicions that not only the young readers we are studying but also the texts they are presented with in tests of reading comprehension could bear improvement.

In short the project's approach can be characterized as (1) DYNAMIC, in the sense that we seek to trace the ideal reader's growing and changing envisionment of the world created by the text; (2) TEXT BASED, in the sense that we seek explanations of the comprehension process by making maximum appeal to the reader's

sensitivity to detailed properties of the text, and (3) EVALUATIVE, not only in the sense that we are able to evaluate the performance of real readers against that of the ideal reader, but also in the sense that we are able to evaluate texts in terms of the cognitive demands they place on their (ideal) reader with respect to prior knowledge, memory, processing strategies, and so on.

1.1.3 Language comprehension in general

We assume that language comprehension in general, and reading comprehension in particular, is a constructive process. By this we mean that the reader takes the information provided by a written text as a blueprint (Spiro 1980, p. 245) for the construction of the text's interpretation, and that any attempt to study the process of reading comprehension necessarily involves an examination of the ways in which a person, knowing certain things and ignorant of others, examines the blueprint and constructs and envisionment of what it is that the designer of the blueprint had in mind. The problems we have taken on are (1) that of devising and defending a way of representing what we take to be the intended interpretations of the kinds of texts we use, (2) that of modelling the process by which the interpretations thus represented can be derived from the text in the case of an interpreter who is in some sense ideally prepared to receive and understand the text (the ideal reader), and (3) that of empirically examining both the process and the product of text interpretation in the case of the text's Real Readers, readers who may be less qualified, less attentive, less motivated, etc. than the Ideal Reader.

We are concerned with observations and speculations about the comprehension process in the face of clearly important but not well understood differences between oral language and written language comprehension. In some respects oral language has to be easier to understand than written language: phrasing, pacing, voice quality, stress, and intonation, to say nothing of the fact that in oral language the participants in the conversation share an immediate context and generally know to what degree they share personal histories, contribute a great deal to the comprehensibility of oral language. In other respects written language has to be easier than oral language, other things being equal: words, phrases and sentences are usually clearly segmented; the signal does not fade; the interpretation process does not have to be paced to the speed of speech; and so on.

The process of understanding written discourse cannot merely be thought of as the process of understanding oral discourse to which the task of word decoding is added. Transcripts of conversations are usually extremely difficult to follow, and passages intended to be read are often impossible to follow when presented by voice. Our interest in the process of understanding written discourse is one which must eventually confront the problem of

distinguishing (1) understanding texts in a given language, (2) understanding that style or register of the language that has been conventionalized for use in written texts, and (3) knowing how to read. The positions we have taken in this study, and the research methodology we have used, do not follow from any specific beliefs about the difference between oral and written communication. In fact, it seems clear to us now that for our purposes we have no obligation to establish or to justify boundaries between those text-interpretation abilities which are tied to language comprehension in general and those which are tied to the process of reading. The materials we are examining are presented to the world as Reading Tests. Our job is to determine as well as we can what is involved in understanding those materials, and then to ask what it is that they ultimately test. Once we have laid out our full understanding of what it takes to understand the passages and what is involved in answering correctly the accompanying questions, we may then begin to ask ourselves how well we think these items test an ability to read as opposed to all of the other skills, background knowledge, and experiences that are required for coping with this material. The question, with our texts, of when it is reading ability that is being tested and when it is some auxiliary skill necessarily keeps coming up; but it is not a question for which we are required to have an answer prepared in advance.

1.2 A sample text: "Nasrudin"

An example of the kind of material the team has been working with is the following passage, taken from the Metropolitan Achievement Tests and intended for third-grade students.

> The carpenter was astonished that such a weird, weak-looking creature as Nasrudin was applying for a job. "Okay, I'll give you a chance," said the doubtful carpenter finally. "Take this ax and chop as much lumber as you can." At dusk Nasrudin returned.
>
> "How many trees have you felled?" questioned the carpenter.
>
> "All the timber in the forest," Nasrudin replied.
>
> Shocked, the carpenter glanced out his window. There were no trees standing on the hillside. Nasrudin has destroyed the entire forest.
>
> "Where did you learn to chop lumber?" asked the astonished carpenter.
>
> "In the Sahara Desert," answered Nasrudin.
>
> "That's ridiculous!" shrieked the carpenter. "There aren't any trees in the desert!"

"There aren't any, NOW," said Nasrudin calmly.

The passage is followed, of course, by a series of test questions. Here is the first question:

1. The carpenter told Nasrudin to

   (a) look for another job
   (b) cut down as many trees as he could
   (c) go back to the Sahara Desert
   (d) plant as many seeds as he could

The children confronting this question, presented as it is immediately after the reading passage, are expected to understand that they are not here being asked to continue the narrative. That is, they have to sense that the story they have just read has been FINISHED, as far as they are concerned, and that they are now being asked to show how well they understood it. Should they by mistake construe their task as one of advancing the narrative beyond the point where Nasrudin made the boastful claim about creating the Sahara Desert, they might find it quite reasonable that the carpenter should advise Nasrudin to look for another job (since he was no longer needed here), tell him to go back to the Sahara Desert (as a kind of "get-out-of-my-life" remark), or indeed order him to plant as many seeds as he could (to make sure that something could get growing on the bared hillsides). The test-takers must first keep in mind the test-taking maxim, that if two answers appear to be equally good both are probably wrong, but must then realize that they are probably NOT being asked to advance the narrative. What they must remember is that in Nasrudin's probationary period, he had been given the ax and told to "chop as much lumber" as he could. They would have to figure out that "chop lumber" is our author's unusual way of saying "cut down trees," (that is, they must sense that "lumber" is the resultative object of "chop"), and they must sense that it was that early conversation between the carpenter and Nasrudin with which the question is concerned.

The second question:

2. How long did it take Nasrudin to complete the job?

   (a) one day
   (b) three days
   (c) thirty days
   (d) three years

In order to answer this question correctly, the children are required to realize that in the sentence, "At dusk, Nasrudin returned," the phrase "at dusk" must refer--for lack of any

contradicting information--to the dusk of that same day; and they must also realize that there is nothing in the story that could back up any answer with the number "three" or "thirty" in it. Those children who are uncommonly sensitive to language will probably wonder what it might mean to "complete the job" under these circumstances, since the only task Nasrudin had been given was to "chop as much lumber" as he could.

The third question:

3. Nasrudin suggested that there were no trees in the Sahara Desert because

(a) trees can't grow in the desert
(b) no one had ever planted any there
(c) they had been destroyed by fire
(d) he had chopped them all down

The answer is that Nasrudin had chopped them all down. This was "suggested," to use the question's word, by Nasrudin's answer, "There aren't any, NOW," said after having given the explanation that the Sahara Desert was the place where he had "learned" to "chop lumber."

The fourth question:

4. After Nasrudin finished work, he

(a) left for the Sahara Desert
(b) told the carpenter what he had done
(c) applied for a new job
(d) yelled at the carpenter

The expected answer is that Nasrudin told the carpenter what he had done. The ordinary scene a reader might construct based on that description, however, is probably a bit different from what we saw in the story, so a certain amount of construing is necessary. The carpenter, it will be recalled, asked Nasrudin, "How many trees have you felled?", to which Nasrudin replied, "All the timber in the forest." This utterance, an elliptical answer to a question which presupposed an understanding of WHAT he had done and speaks only to the question of HOW MUCH he had done, has to be construed as an instance of Nasrudin telling the carpenter what he had done. The tempting possibility that the correct answer is "yelled at the carpenter" is presumably introduced to take advantage of the printer's decision to put the word "NOW" in capital letters and to end the sentence with an exclamation point: "There aren't any, NOW!." The capitalized word suggests shouting, so what is being cleverly tested with this foil is the child's ability to notice that this natural conclusion must get corrected by the next three words, "said Nasrudin calmly."

The fifth question:

5. The carpenter had not expected that Nasrudin

(a) had ever seen the Sahara Desert
(b) really needed a job
(c) would be so rude
(d) could do the job so quickly

In this item the pluperfect form, "had not expected," plays an important role. The sentence has to be situated in the text at some time point where it serves an explaining role. The text reveals that something was unexpected at that place where it shows the carpenter surprised. Being "shocked" is an extreme form of being surprised. The text describes the carpenter as "shocked," through a grammatical device whose function may not be transparent to most third-grade readers, and backs this up by showing that the carpenter spoke from then on only in sentences ending in exclamation points. He had been shocked when he learned--and hence he had not expected that it had been possible--that at the end of the first day on the job Nasrudin had conquered the entire forest. Again, describing Nasrudin's accomplishment as doing "the job" quickly, in the face of the actual instructions he had been given, will seem strange to the linguistically sensitive.

1.3 The purpose and nature of our interviews

It is of concern to our project how young readers construct an understanding of reading passages of the kind we have just seen, and how well that understanding can be appealed to in finding "best answers" to test questions about those passages.

There are two main phases in the work of our project. The first is the Analysis phase. In it we analyze reading-test items, both the passages and the test questions, in such a way as to be able to isolate and describe the kinds of background knowledge and the kinds of interpreting and integrating skills which a reader must bring to the passages in order to get out of them what their creators intended. In the second phase we seek to find out, by interviewing our young subjects, whether in fact they had that knowledge and those skills. We intend the results of this second phase to be the development of a technique for evaluating the intelligibility of reading passages and a method of criticizing test passages and their associated questions.

We began by analyzing test passages and devising a system of annotations for them which were intended to serve three purposes: it represented our view of the comprehension process of the ideal reader of the text; it provided the material for our choice of probe questions to be used in the interviews with the children; and it gave us a framework or checklist against which we are able to evaluate the children's performance with the items. The second

part of our work, then, involved close observation of the children's experiences with the items, in which we made use of free retellings of the substance of the passages, interviews and metacognitive probes, looking for the presence or absence of the kinds of knowledge and skills which the passages require of their ideal readers.

In exploring the real readers' experience of the texts, we presented the texts to them one segment at a time. The general method we used worked like this: we would show a child the first segment, then ask our first batch of questions; then we would show the child the second segment, asking our second batch of questions; and so on. As each increment was added we made sure that the entire preceding portion of the text was still available for scanning, re-reading and consultation, and that the child was free to go over that part of the text at will.

The method required us to make arbitrary (not random, of course) choices of the segments which were to be examined one at a time. To the extent that we chose unwisely, our results fail to reveal the actual dynamics of the texts' processing for real readers. Our goal was to find out to what extent the children had the requisite knowledge to integrate new pieces of information from a text with the old, so our inclination was to choose segments that were probably smaller than the ones real readers might naturally choose; that inclination, however, was balanced by the pressures of time in the interviewing, since with the very short segments the interview became long and boring. We generally chose phrases that we felt could easily be expressed in a single breath-group. Just in case variable chunking is one of the sources of children's miscomprehension, our method does not bring that out. Obviously, research different from ours is necessary for determining the consequences of skillful and unskillful text chunking.

The sheer mechanics of our kind of presenting and interviewing proved to be interestingly difficult. Since some of these difficulties shed light on the process of text understanding, a discussion of them merits some place in this report.

At first our method was to decide on a particular segmentation of a text, type it out with each segment on a separate line, and have the subjects slide a piece of cardboard down over the text one line at a time during the interviewing process. This method introduced two major difficulties. First, as the card got lower on the page, the exposed piece of text at the top ended up looking more like verse than prose, and that, we felt, was affecting some people's interpretations of what they were reading. Second, the method did not make it obvious to the subject how much of the text was left, and we felt that readers might use different strategies for interpreting a sentence if they thought it was the closing sentence in a text than if they thought more was coming. This uncertainty made the experience quite unlike normal reading;

in normal reading, we almost always know how close we are to the end, and that knowledge plays a large part in shaping our expectations and putting our interpretative faculties to work.

Our current method is more successful, but it too has problems. We type the text on a sheet of paper, in the normal way, double-spaced. If the text has been given N segmentations, we make N+1 xerographed copies of it, and construct a booklet. Page 1 of the booklet has the entire text blocked out with a marking pen. Page 2 has everything blocked out but the first segment. Page 3 has everything blocked out but the first two segments. And so on. With this method, then, as the text gets exposed, it looks like ordinary printed prose, and it is very clear to the subject how much of the passage is left before the end is reached. (The passages we use all fit on one page.)

1.3.1 A Model interview

The interview works something like this. The child who turns the first page sees

Once upon a time XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
etc.

exposed at the top of the second page. The interviewer asks something like, "What can you tell me about this passage so far?". The subject answers, "Well, it's going to be a story. Most likely a fairy tale." "How do you know that?" "You only say 'Once upon a time' when you're telling a fairy tale." "Do you have any guesses about what we're going to read when we get to see more of the story?" "Well, maybe something about a poor old lady who lived in the woods, or maybe a rich king with a beautiful daughter. I don't know."

The child then turns the page to expose the next increment, and sees:

Once upon a time there was a rich king
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
etc.

The interviewer says, "Say, you were right, weren't you? The sentence isn't finished yet, is it? Do you want to stick to your guess about the beautiful daughter?" "Yeah."

The child turns the page and sees:

Once upon a time there was a rich king
who had three sons. XXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
etc.

And so on through the text.

The new method of presentation solved some of our problems in exploring the reading experiences of real readers, but there remain a number of fairly serious problems connected with the piecemeal presentation no matter how it is adapted. One difficulty with our method comes from the fact that the talk produced by the interviewer was itself a text which the children had the right to believe had a point and a direction. Imagine leading children through a text about people who live in trees in which that information is revealed as a surprise after some time has elapsed. If, prior to the revelation, we were to ask questions like, "Do you think these people live in houses like everyone else?", in order to be able to predict whether our young subjects were going to be surprised when they found out the truth, the children would quickly catch on to our purposes. If we tried to balance this by asking dozens of questions about normal readers' default assumptions ("Do you think they grow hair on their teeth?" - "Do you think they sleep at night?"), our questions would be intolerably distracting.

The interview questions were sometimes unavoidably distracting. For example, in the text described below under the name "Bronco Buster" the noun compound "bronco buster" was one of the items which was unfamiliar to many third graders. In order to find out whether they at least had the ability to recognize the compound word structure "Noun Verb-ER" an interview question was introduced that made use of a compound we felt sure the chilren had never before encountered. We asked, "What would you think if something or somebody was described as a PIG WASHER?". Every child knew that this had to be something or someone that washed pigs. The follow-up question, designed to discover whether the children sensed the structure of the compound, was, "Now what do you think a bronco buster is?". Two of the children in our sample answered, "Someone who washes broncos?"! One may imagine avoiding this distraction by waiting until the child had finished reading the text before asking about "bronco buster." But of course if one selects this option, one has no way of telling those children who understand "bronco-buster" on first encounter from those who were able to figure out what it must mean only by reading the entire text.

A further problem produced by the interviewer's own text shows up when we explored the reasons for the choice of particular test question answers. With some young children, the question, "Why did you choose this answer?" suggests powerfully that the

answer was wrong.

One source of difficulty, then, resides in the character of the conversation with the interviewer. Another is the pacing. A paragraph that takes less than a minute to read in normal circumstances can take half an hour to go through in the segment-by-segment presentation. The interview brings so many things into the subjects' consciousness that the simple thread that was our text could get completely lost. Using the method with adult readers on adult texts, we learned that passages that are humorous when read all at once are not humorous when given out piece by piece. It is not merely that the passages do not seem funny; sometimes their humorous intent is not even discerned. And with young readers the factor of interest may work against us both ways. If the story is interesting, the reader may want to get to the end to see how things turn out, and become impatient with so many boring questions. If the story is uninteresting, the reader may want to get the whole experience over with as soon as possible, and becomes impatient with so many boring questions. A lot depends, in short, on the warmth and charm of the interviewers, and on whatever rewards children may feel in knowing that an adult is paying very close attention to their words and thoughts.

A partial control for the unnaturalness of the piecemeal presentation method was established by having half of the children in the study read the whole text first, all at once, at their own pace. These children were then asked to retell the passage in their own words, and were then asked to go through the text once more, with the interviewer, in the segment-by-segment way. The questions for this group had to be differently formulated; they included things like "What were you thinking when you read this?" - "Do you remember what happened next?" - and so on.

There were still other problems with the method, involving the interpretations we could give to children's responses. One of the first transcripts we examined with care made us cautious of the answer "I don't know." Jasper Meadow had read

If a bronco buster wants to win a rodeo
contest XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

and had reported complete ignorance of the meanings of the words "bronco buster" and "rodeo". But then, in answer to the interviewer's question, "Do you have any idea what we're going to read about in the rest of the passage?", Jasper said, "People riding on horses?".

1.3.2 The problem of genre in reading generally and in reading tests in particular

In later sections we will discuss the fact that the schemata we attribute to the reader as his principal conceptual tools in

construing a text may be divided into three classes: content schemata comprise our knowledge of the world; text schemata comprise our knowledge of grammar and of how a text larger than a single sentence is built; genre schemata comprise our information about certain text types that unite specific text forms with associated communicative purposes and also usually with certain kinds of subject matters. Examples of textual genres are lyric poems, assembly instructions, adventure stories, sermons, and so on. Our reason for considering the topic of genre in the present section is that reading tests pose special problems with respect to genre.

First among these problems is that while each reading test item pretends to belong to a normal genre such as a narrative about a boy and his horse, an expository paragraph about bananas or industrial dust, or a folktale about a clever trickster, this normal genre is always embedded in a constant matrix genre: the reading test item genre. We have mentioned that each normal genre of writing usually corresponds to a more or less specifiable communicational purpose, for example expository prose to the communication of structured information, narrative to entertainment and/or the communication of moral points, jokes to entertainment, and so on. Along with these communicational purposes, there are also predictable textual structures, such as topic sentences and conclusions in expository prose, settings, character introductions, conflicts, resolutions, etc., in narratives, build-ups and punch lines in jokes. In order to be able to construe a text correctly and to experience the text in the way its author intends, a reader must be able to recognize the genre of a text, form the appropriate expectations and draw the corresponding conclusions. This is as true of the reader of the pseudo genres of reading test items as of true genres. Thus, in a reading test item that contains a story about a wild pony who gets his hoof trapped between some rocks and is about to be attacked by a hungry wildcat, when the reader is suddenly told that the pony's mother is a "big brave horse," the IR has already interpreted the text so far as an adventure narrative with the little pony as the protagonist and the wildcat as the antagonist and potential peril and is able to predict that the mommy will come to fight off the wildcat, providing a resolution to the problem that besets our hero. That is, the ideal reader of a reading test item is able to recognize the embedded genre and make appropriate calculations. But the reader of a genre embedded in a reading test item must not make the mistake of thinking that he or she is actually reading a text of that genre, because such a reader would then mistake the communicative purpose aspect of the genre with that of the text with which he or she is actually confronted. For example, if the embedded genre is a story, the reader had better not read for enjoyment--ignoring detail that doesn't advance the narrative--as such detail is just the thing most likely to be required for answering the questions on a reading test. Thus, from the point of view of genre, the reader of a reading test item must always be carrying out calculations at two levels. At the lower level he or she must figure out what kind of genre is being imitated by this

reading test item and make the appropriate assumptions, calculations, and conclusions. But at the higher level the same reader must always bear in mind that this is really a reading test item, that he or she must absorb the information in such a way as to be able to answer questions with single word answers chosen from sets of four, that this is reading for total short-term recall, not for fitting this information into what he or she already knows about the world, nor for enjoyment, nor for discovering the kind of information the text contains in case he or she ever needs to look it up in the library, and so on. In short, the peculiar pragmatics of the reading test situation require that the reader always deal with genre at two different levels simultaneously. This is one respect in which adequate performance on reading tests requires skills different from those required for normal reading and suggests that scores on reading tests reflect to an unknown degree the extent to which the pupil possess these test-specific skills.

The situation in which the child confronted with a reading comprehension test item finds itself is in fact worse than the foregoing description would suggest. In the situation we have just described, the child confronts a text that has the form of, say, a narrative, but must be read in a different way (for different purposes) than a narrative is read. Actually, the textual form of a reading test item is frequently not a competent realization of the genre the item is pretending to emted. The reasons for this appear to fall into two main classes, one shared with other reading materials given to school children and the other special to reading tests. The former set of reasons includes the avoidance of copyright infringement and the necessity to meet readability formulas. Thus the texts of reading comprehension tests, like other texts presented to children, are frequently deformed by the omission of sentential connectives or in other arbitrary ways to produce a smaller word/sentence ratio or to obscure the connection with the texts from which they are adapted.

The latter set of reasons seems to be peculiar to reading tests and perhaps one or two other pseudo-genres that share with reading test items the property of presenting texts that are abnormally short. (Examples in scientific papers about AI story understanding programs often suffer the same deficits.) For example, such texts often introduce a participant with a type of noun phrase used in normal texts only to refer to a participant who has already been introduced. Such devices are proper names, definite descriptions and pronouns. Thus in place of the normally expectable "Once there was a little pony named Silver," we find in a reading test text the opening sentence "Silver was a little pony." Although one may suppose that this habit of beginning texts with grammatical forms normally encountered only after a text is

underway, generally motivated by the excessive brevity of these text, in this example, as in many, very little economy of length is in fact achieved and one is inclined to conclude that the employment of these peculiar devices has achieved a kind of functional autonomy as a mark of the genre. This is another case in which the reader of reading test items has to do more and different work than the reader of normal texts. In general the embedded genres of reading test items are truncated or schematic; we can recognize this text as a caricature of a story about a pony or a caricature of an informative article about the first phonograph. (The "story" lacks any denouement; the "informative" article presupposes that the reader already has the visual image it is pretending to convey.) Thus the successful reader of these texts has to know not only the rhetorical devices that signal the genres being imitated but also the special rhetorical devices that signal the caricatures of these genres that are to be found in reading tests.

Success in reading these texts is of course judged by the answers given to the questions that follow the text. The questions themselves and the answering strategies they require are discussed in detail below. With reference specifically to reading tests as a genre, it is important to realize that the young reader must appreciate that these questions also play a very special role within this complex and idiosyncratic textual genre.

2. The Ideal Reader Analysis and its Use in the Interviews

Our analysis of a text in terms of the Ideal Reader's processing of that text and our interview with real young readers are linked by a three-column document which gives (1) our textual analysis of the items according to the Ideal Reader's processing, (2) the textual and theoretical justification of each judgment attributed to the Ideal Reader under (1), and (3) the question(s) to be posed to the real reader aimted at discovering the extent to which the real reader is processing the text at each point in the same way as stipulated for the Ideal Reader. The general idea of this three column document will be briefly illustrated here. Then we will discuss the details of the Ideal Reader analysis and the principles of textual interpretation that motivate and justify it in greater detail. One of the texts we have dealt with, we have named "Bronco Buster". As reproduced below, each portion of "Bronco Buster" between slashes corresponds to a segment presented as a whole to the young readers in our interviews.

> /If a bronco buster wants to win a rodeo contest,/ he must observe the contest rules./ One of these rules is that the rider must keep one hand in the air./ A rider who does not do this is disqualified./

The first page of our three-column analysis and question-prompt document for "Bronco Buster" is reproduced below. This page is devoted to, and does not exhaust, the analysis of the first segment of the text.

> If a bronco buster wants to win a rodeo contest,

BRONCO BUSTER
January 29, 1981

1] *If a bronco buster wants to win a rodeo contest,*
LO(I) If-clause, present tense, indefinite articles

1. Sch (Co) CONTEST PARTICIPANTS, DESIRE TO WIN, COMPETITION, RULES, CRITERIA FOR DECIDING ON WINNERS,...	1. 'win' & 'contest'	1. "Can you tell me what happens in a contest?"
2. Sch (Co) RODEO CONTESTANTS, AUDIENCE, EVENTS: RIDING, ROPING, TYING, ...	2. 'rodeo'	2. "Do you know what a rodeo is?" "What kinds of things happen in a rodeo?"
3. SCHEMATIC LINKS a RODEO presents a number of CONTESTS		
4. Sch (Tx) N+N COMPOUND N^1 identifies a type of N^2	4. 'rodeo contest'	4. "Would you say a rodeo contest is a kind of contest or a kind of rodeo?"
5. Sch (Co) BRONCO BUSTING HORSE: WHEN WILD WILL TRY TO THROW RIDER RIDER: RIDES WITHOUT GETTING THROWN, BY BREAKING WILL OF HORSE RENDERS HORSE TRAINABLE	5. 'bronco buster'	5. "Do you know what bronco busting is?" "Do you know why they do it?"
6. SCHEMATIC LINKS BRONCO BUSTING is one of the kinds of events in a RODEO		
7. Sch (Tx) N+V-er COMPOUND cpd designates someone who V's Ns	7. 'bronco buster'	7. ("What would you say somebody who is a pig-washer does? What about a mailbox painter?")
8. Sch (Tx) CONDITIONAL SENTENCE ANTECEDENT, CONSEQUENT	8. 'if'	8.
9. Hyp↓ (Gn) GENRE = EXPOS PROSE	9. doesn't look like a poem, or a letter, or a story	9. "What kind of a 'story' do you think this is? Do you think it's going to be the kind of story you might read in a story book? ... (ad lib) "
10. Hyp↓ (Th) THEME = WINNING RODEO CONTEST	10.	10. "What do you think this is going to be about?"

The document first reproduces the initial segment of text just mentioned. The notation just below this calls attention to particular aspects of the grammar of the text that we deem either crucial to the IR's correct construal of the text or potentially to pose special problems for the young reader, or both. In this case, the items listed refer to the facts that (1) the segment presented is an if -clause, (2) the main verb is present tense, and (3) both of the noun phrases are indefinite. In lines 11 and 12 (not shown on this page of the document) the IR is called upon to put together these three facts about the grammar of the segment in order to realize that the "bronco buster" in question is a generic bronco buster and this is the beginning of a condition sentence that will go on to say something about what bronco busters in general have to do if they want to win rodeo contests. Most saliently, the IR must realize right here that this is not a story about a particular bronco buster but probably some kind of didactic texts about constraints on being a successful bronco buster in the abstract. (We may note in passing that it is not clear that school children ever encounter exactly this genre of text other than as an embedded genre in a reading test text.)

Line 1, in the left column, notes that the segment evokes the content (world knowledge) schema of a contest and lists some elements of that which will be evoked in the mind of the IR. In the middle column are given the lexical items that conduce to the evocation of this schema. In the right column we present the question that is intended to elicit what the actual reader being interviewed knows about this schema.

Line 2 deals in an analogous way with the rodeo schema.

Skipping to line 4, the left (analysis) column notes the textual schema that must be evoked if the IR is to properly process the N1 + N2 nominal compound rodeo contest, in particular that given the knowledge of what a rodeo is and of what a contest is the IR will know that a rodeo contest is the particular kind of contest that occurs in connection with rodeos. The middle column gives the warrant for this: merely the presence of the compound word rodeo contest in the text. The question listed in the third column of this line deals with only part of this problem for the IR, questioning the knowledge that in an N1 + N2 compound, the category named is a subcategory of that named by N2.

Skipping to line 7, the same textual problem is taken up in greater detail. Here we seek the more detailed knowledge that a N1 + N2 compound like this, which has the shape N1 + V-er, means 'someone who V's N1s', that is, in this case, that a bronco buster is someone who busts broncos. Hence the question in column three, which seeks to elicit this abstract knowledge by querying a different instantiation of the same word-forming pattern, to see if the child knows that a pigwasher is someone who washes pigs.

The point of this kind of detailed inquiry is that there is

more to knowing what is going on in the reader of this passage with respect to the item bronco buster than simply the answer to the question whether or not the reader of this compound successfully constructs the image of the appropriate category. One reader may know the word bronco buster and what it denotes as a single unit; another may never have encountered the expression bronco buster before, but may be able to figure out all or some of its meaning in any of a wide variety of ways. To give but a simple example of the latter kind of reader, suppose this reader knows that the word bust is a colloquial variant of the word break, that the word break has a special use in which it may be substituted for the more generic verb tame just when the object of the taming is a wild horse, and that bronco is a word naming a particular kind of wild horse. This reader might then be able to put all these things together and decide that a bronco buster could be someone who tames wild horses.

In fact there are several other possibilities of this kind that combine varying mixes of world knowledge, vocabulary knowledge and knowledge of text and grammatical structure; these will enable different kinds of readers to figure out some, all, or none of the significance of the compound bronco buster. When our young readers fail, we think it is important to know just how they fail, i.e., just which items of textual, lexical, or content schemata they lack, or whether what is lacking is not the schemata themselves, but the ability to make the calculations entailed in their combination. Curiously, in actual rodeos, the activity is normally called bronc riding and seldom or never bronco busting, nor are rodeo bronc riders normally called bronco busters. Hence the young reader who knows a lot about rodeos and their associated nomenclature, but who is unable to do some kind of grammatico-etymological analysis of the peculiar terminology used here--which properly refers to a utilitarian activity on ranches, not the entertainment activity at rodeos that is modeled roughly on it--is likely to have difficulty with this text. Of course, professional rodeo performers may or may not also be tamers of wild horses, probably most often not. Whether tests of reading comprehension should require this kind of knowledge and calculation is doubtful.

Line 9 illustrates a different kind of operation than the ones we have considered so far. Up to now we have discussed the evocation of schemata (either grammatical, content or textual schemata). In line 9 we encounter an expectation formed by the IR. (The notation "Hyp" for "Hypothesis" with a downward arrow is intended to be mnemonic for "expectation": that is, a hypothesis about the future of the text.) The justification given in the middle column may be judged rather thin. It gives neither a specific reference to one or more items of the text that evoke this response, nor does it mention a general principle of text construal. Certainly the discussion above about the combination of the grammatical elements if- clause, present tense, and indefinite noun phrases is part of the story. Sometimes, however, we find it necessary to attribute to the IR of a certain text the ability to

make a particular judgment (in this case an expectation) without our being able to give fully explicit warrant for it. Such judgments are useful nevertheless, both in guiding the interviews with young readers and in sharpening the kind of questions we ask ourselves in developing a more complete theory of the IR.

In section 2.1 we will consider in greater detail the mechanics of IR analysis that goes into column 1. For example, we will discuss other sorts of processing operations that contrast with the evocation of schemata and the formation of expectations. And we will discuss more systematically the levels of processing such as content, text, grammar, and genre. For the moment we want our reader to bear in mind that this IR textual analysis forms one part of the three-part layout we use as the basis of our interviews.

2.1 Column One: Notations of the IR's processing of the text

We distinguish three domains for the Ideal Reader's speculations, puzzlements, and conclusions: the domain of CONTENT, by which we have in mind the properties and events in the world of the text; the domain of TEXT, with its schemata of grammatical structure and text structure; and the domain of GENRE, where we have in mind those structures of expectation that come with knowing that one is dealing with a folk tale, a detective story, an obituary, or a reading test.

Genre schemata arise from structured expectations created by familiarity with particular genres. If we read in a folk tale that the king's two older sons have both failed to slay the dragon, we are filled with hope when we learn that the king's third and youngest son has set out to try his hand. Were the story to end with the third son being slain by the dragon, we would feel that we had just been exposed to a new and cynical derivative genre, not that we were wrong in forming the expectations we had formed. When we read questions in a reading test, we know that we are not being asked to figure out a clever way to finish the passage we have just been reading, but rather that we're being asked what we remember, or what we can now figure out, about what the passage told us. Some genres are keyed to special lexical items and phrases to a greater degree than others. For example, fairy tales are full of words like damsel, sally forth, and so on, as well as formulas such as once upon a time. Similarly, special uses of ordinary words, e.g., land, in the sense of country or locality (as in a far-away land, the land of the green fountain) may be restricted to single genres or small families of genres. Other clues to genre may be considerably more abstract and structural. For example narratives have protagonists, settings, problems, resolutions, and denouements, none of which concepts are automatically instantiated by particular surface lexical items or phrases.

As we see it, the main dynamic aspect of the reading

experience is that of constructing and revising an Envisionment of the "world of the text," some coherent "image" or understanding of the states of affairs that exist in the set of possible worlds compatible with the language of the text. We do not intend the word envisionment to suggest too strongly the visual aspect of a text world, but at the same time, we recognize that in the kinds of texts we are dealing with, the visual aspects often predominate.

The envisionment is the representation in the reader's mind of the content of the text. The envisionment grows, and sometimes changes, as the reader (or hearer of a monologue) progresses through the text.[4] According to this view, the ideal reader, as it reads a text, not only updates and supplements its envisionment of the world the text is describing, but also--in the service of building this envisionment--formulates hypotheses, asks questions, notes evidence and in general, accomplishes a variety of processing operations.[5]

The ideal reader does not invest in every part of its envisionment the same degree of confidence. The ideal reader does a lot of reading between the lines. But the ideal reader also knows that it must place less trust in those parts of its envisionment that arise from reading between the lines than in those parts that come from reading the lines themselves. The ideal reader in fact distinguishes many levels of confidence within the envisionment. Consider the two text fragments:

(2) The Orioles' shortstop bunted the ball right to the first baseman, who grabbed it and tagged the batter out.

(3) The Orioles' shortstop threw the ball right to the first baseman, who grabbed it and tagged the batter out.

Let us catalogue several inferences the ideal reader will draw from these texts, that is, several items that ought to become

---

[4] Reading is probably easier to model in this way than hearing is because reading lacks the interactive phenomena, such as negotiation of interpretive context, back channel signalling, and the like, that are the bread and butter of the conversational analyst. Nevertheless, reading is hard enough. We are of course dealing here specifically with reading, but to the extent the model we are developing is correct, it may be useful in the understanding of the text semantics of spoken language as well. Written language also differs from spoken language in other ways, such as the imperfect substitution of punctuation for intonation, stress, and prosody, as discussed in section 1.

[5] Fillmore and Kay 1980: 22-49.

part of the ideal reader's envisionment of these texts. In particular we want to look at relations of coreference between pairs of noun phrases in the two texts. First of all, we note that in both (2) and (3) the pairs of noun phrases (the ball, it) and (the first baseman, who) are coreferential, which facts are summarized in Table 1.

Text (2)	Text (3)
(a) the ball - it	(a) the ball - it
(b) the first baseman - who	(b) the first baseman - who

Table 1

Facts (a) and (b) reflect directly the grammar of the texts, and nothing more. This may be seen by considering text (4), which is parallel in grammatical structure to both texts (2) and (3) (which themselves are grammatically parallel, differing only in the alternation between the lexical items bunt and throw), but in which nonsense words are substituted for the content items.

(4) The Wimbats' glurb slunked the wint to the girfman, who critched it and ...

The reader will appreciate that the grammar of (4) determines that the relations of coreference must be as shown in (5), which accord with those of Table 1.

(5) The Wimbats' glurb-i slunked the wint-j to the girfman-k, who-k critched it-j, and ...

Thus inferences (a) and (b) are not dependent on any knowledge or skill the reader may possess beyond his or her knowledge of English grammar. In particular, no knowledge of the game of baseball is required. Also, and more importantly, it is impossible to imagine a continuation of text (2) or (3) that could overthrow these inferences. For example, if we load the content to make conflicting inferences regarding coreferences more plausible, the result is not a switch of perceived relations of coreference but a judgment of incoherence. Thus in (6) we rig the content to try to get the relative pronoun who to be coreferential, not with the immediately preceding noun phrase, but with the initial noun phrase; the attempt fails.

(6) *The batter-i grounded to the shortstop-j, who-i was thrown out easily.

In short, the reader need know nothing about baseball and, in particulr, need know nothing about the meanings of the words shortstop, bunt, first baseman, and so on, to infer the facts given in Table 1. Further, there is no way in which subsequent

developments of the texts (2) or (3) can induce the English speaking reader to change his mind about the inferences (a,b) of Table 1.

Inferences (a) and (b) were the same for texts (2) and (3). There are inferences of the same general kind in which the two texts differ. In particular, in text (2) the Orioles' shortstop is the batter while in text (3) the Orioles' shortstop is not the batter. Table 2 adds these inferences to those of Table 1.

Text (2)	Text (3)
(a) the ball = it	(a) the ball = it
(b) who = the first baseman	(b) who = the first baseman
(c) the Orioles' shortstop = the batter	(c) the Orioles' shortstop ≠ the batter

Table 2

But now we note that these new inferences (c) are in part dependent on schematic knowledge of the game of baseball. In particular, they are dependent on the knowledge that shortstop is the name of a fielding position, that batters may be referred to by their fielding positions, that bunting is something that batters (not fielders) do, that throwing is something that fielders (not batters) do, that the Orioles is the name (or could be the name) of a baseball team, and so on (the list is not complete). The reader who has absolutely no knowledge of baseball will notice that he or she was unable to derive inferences (c).[6]

It is not impossible to imagine a continuation of the texts (2) and (3) that could vitiate inferences (c). We can imagine a game that is generally like baseball except the offensive player is not equipped with a bat but rather attempts to catch the pitch and throw it out into the playing field. The defensive players, on the other hand, are equipped with bats and are required to hit or bunt the ball to the basemen to put the batters out. A skilled writer might be able to build a text that took (2) or (3) as the

---

[6] Readers who doubt that inferences (c) require some knowledge of baseball should consult an acquaintance who lacks all such knowledge. To make the point I am making here, it has been necessary to select an example from a domain about which knowledge is widespread but not ubiquitous among English speakers Those who lack knowledge of baseball at the requisite level will have either to take my word for it or consult someone who has such knowledge. No affront to either knowers or non-knowers of baseball is intended.

initial fragment and develop it along these lines so that the reader would ultimately reject his or her initial inference of type (c) and arrive at a final envisionment of a game of the sort just suggested. A text that began with fragments (2) or (3) and caused us to overthrow inferences (c) might seem far-fetched, perhaps some kind of literary tour de force. Nevertheless such a text is possible in English. In this important respect inferences of type (c) differ from those like (a) and (b), although one suspects that with regard to actual texts and actual readers the analytical distinction is unlikely to make much practical difference. Nonetheless, in thus distinguishing inferences (a,b) from (c), we have detected what we may call two levels of envisionment: the (a, b) level being absolutely ordained by the text and the (c) level being strongly--though not absolutely--determined by the text. The point, which we shall now pursue, is that items of the envisionment may be warranted by the text to varying degrees and therefore to varying degrees cancellable by further development of the text without discomfort to the ideal reader.

Consider now the inferences, obvious enough to those fairly familiar with baseball, that in text (2) the first baseman is not an Oriole while in text (3) the first baseman is an Oriole. These inferences require grammatical knowledge and the knowledge of baseball (and the lexicon of baseball) of the kinds already considered. In addition they are based on the world knowledge that the players on one team of a given game normally belong to one ball club while the players on the other team belong to another ball club. That is, we normally think of baseball games which are known to involve players on regularly organized, league ball clubs as being regular league games, in which the players represent the teams of which they are regular members. There are, however, regularly occurring, if less frequent, events in American culture in which these conditions do not hold, for example, all-star games. The rosters of all-star teams are necessarily composed of players from a variety of teams. If the game being described in text (2) or (3) turns out to be an all-star game, then the text provides no warrant for the inferences just mentioned, which are given as (d) in Table 3.[7]

[7] There are other plausible continuations of texts (2) and (3) in which inferences (d) need not--in some cases cannot--hold; examples include sandlot games and intra-squad games.

Text (2)	Text (3)
(a) the ball = it	(a) the ball = it
(b) who = the first baseman	(b) who = the first baseman
(c) the Orioles' shortstop = the batter	(c) the Orioles' shortstop ≠ the batter
(d) the first baseman ≠ an Oriole	(d) the first baseman = an Oriole

Table 3

Inferences (d), although perfectly natural to draw from these texts are, as we have just seen, relatively easily suspendable under culturally non-farfetched, contextual assumptions; this is the same as saying that they are easily suspendable under relatively non-farfetched continuations of the same texts. We want to say, therefore, that inferences of types (a, b), (c) and (d) belong to increasingly "higher" levels of envisionment, where the lower the level the more direct and absolute the textual warrant for the inference and the higher the level of envisionment the more contingent and revisable the inference is. Roughly speaking, if a text requires revision of something in the envisionment at the lowest level we say that the text is inconsistent, self-contradictory, or incoherent; if the text requires revision of something at a "medium" level of envisionment we are inclined to say that the text is surprising; if we find that the text requires revision of something we have at a very high level of envisionment we are inclined to attribute no particular property to the text. Rather, we may register surprise that we as readers have populated our envisionment of the text with some item that lacks textual warrant.

For example, suppose the reader envisioned the Orioles' shortstop as right handed. This would be reasonable as, not only are the majority of people right handed, but the exigencies of baseball play are such that very few if any professional shortstops are left-handed.[8] The reader who envisioned the Orioles' shortstop as right handed would have a right to be mildly surprised if it turned out later in the text that he wasn't, and the entitlement to surprise would increase with the length of text that intervened between the introduction of the shortstop and the revelation of his left-handedness. Compared to the (d) inferences about which team the first baseman is on, this inference about

[8] Left handedness is common at other positions; e.g., pitcher and first baseman.

handedness is less directly warranted by the text and more by general background knowledge of baseball; it also seems less likely to figure in coherence relations of the kind we will discuss under the parsimony principle below. For these reasons, we may with fair confidence add it to our table as representing a still more tenuous level of inference.

Text (2)	Text (3)
(a) the ball = it	(a) the ball = it
(b) who = the first baseman	(b) who = the first baseman
(c) the Orioles' shortstop = the batter	(c) the Orioles' shortstop ≠ the batter
(d) the first baseman ≠ an Oriole	(d) the first baseman = an Oriole
(e) The Orioles' shortstop is right handed	(e) The Orioles' shortstop is right handed

Table 4

At a level of inference or imagination more tenuous yet, suppose some reader forms an image of the shortstop with a mustache. If it turns out later that the shortstop is clean shaven, any reader who experiences surprise is not, we think, entitled to account for this experience by saying that the text is surprising. Hence we do not add this kind of inference to the ideal reader's envisionment at any level.

The ideal reader will thus make inferences regarding the content world of the text at a variety of levels. As we have noted, the reader who cannot or will not read between the lines is not a good reader, and the reader who is unable or unwilling to distinguish what he or she has been told from what he or she has inferred or imagined is also less than ideal. The latter ability, to distinguish what is directly warranted by the text from that which is less directly warranted (even to the point of what is not warranted at all by the text but is merely contributed idiosyncratically to the envisionment), is most obviously necessary to the reader of informational texts, or, more exactly, to the reader who approaches a text for the purpose of assimilating the information it contains.[8]

[8] Some texts--for example, assembly instructions--seem inherently informational while others--for example, poems--seem inherently intended to arouse aesthetic experiences rather than t impart objective information. It is, however, possible and perfectly normal to approach texts originally intended to be taken one way in

Our notation keeps track of the ideal reader's growing envisionment as it progresses through the text. It also records the various processes that we take to underlie the IR's achievement of its envisionment of the text. This report will not present a handbook for the use of our notation (for an earlier version, see Fillmore and Kay (1980)), but some words about its content and the motivation for that content are in order here.

With respect to levels of envisionment, although recognizing that theoretically one should probably recognize a continuum of degrees of certitude, we have found that for the practical analysis of short texts three levels suffice. These we notate as E-0, the absolute level strictly ordained by what the text says; E-1, highly probable inferences from the text which will cause great surprise if they are contradicted by non-immediate future developments and part of whose motivation is the creation of coherence in the envisionment; and E-2, likely inferences which are not required for achievement of textual coherence but which follow from world knowledge about the kinds of things the text is describing.

The envisionment is populated by images and propositions. Developments in the text can move an item of the envisionment to a higher or lower level, may introduce a new item or may expel an existing item. Much of the IR's processing has to do directly with Questions, Hypotheses and Evaluations (of hypotheses) regarding what should be in the envisionment at what level. These three types of processes are discussed below.

We consider first, however, the major process of Evocation. Items of textual material may evoke schemata at any of three levels: content, genre, or text. A content schema relevant to the "Bronco Buster" text is that of a RODEO. It is evoked by the word rodeo. Content schemata are often evoked by words that name their parts; e.g., waiter or menu will evoke RESTAURANT. It is not so easy to name a relevant genre schema for this text, although, as discussed briefly above, our notation shows how the IR is concerned to assign this text to some genre and is at least able to reject certain genres. The co-presence of an if- clause, indefinite reference and present tense in the first segment rules out many genres, notably narratives. Our prior discussion of the IR's calculations with regard to the nominal compound bronco buster illustrated the evocation of two textual schemata: nominal compounds of the form N-1 + N-2 and the (non-conflicting) schema of nominal compounds of the form N-1 + V-er. That discussion also

another. Such is the case, for example, when the literature student reads a play or poem to learn its content and structure rather than to experience its aesthetic values. There is, of course, much more to the subject of what functions a text may fulfill than is suggested by the simple contrast "impart objective information" vs. "arouse aesthetic experience."

illustrated the phenomenon of interaction of schemata at different levels, which we will have occasion to say more about later.

The classification "text schema" takes in a wide spectrum of phenomena. Strict grammatical knowledge instantiates this category as does knowledge relating grammatical form to world knowledge of the sort just discussed in connection with the nominal compound bronco buster. Lexicalized rhetorical attitudes, represented by certain uses of words such as nevertheless, anyway, moreover, well, and so on also serve to evoke textual schemata. A final example of this non-exhaustive listing of types of textual schemata involves a range of phenomena (roughly coextensive with that considered by Grice in his Logic and Conversation ( )). These involve the addressee schematizing what he or she reads (or hears) and his or her purposes and resources. There are certain inferences that we draw, not from the information that is communicated to us, but rather from the fact that it was purposely communicated to us by another presumably rational being. Thus, if we observe a man from the left and note that the visible eye is blue, we are likely to conclude that his right eye is also blue; but if we are told of a man that his left eye is blue, we are likely to conclude that his right eye is not blue. "Textual schemata," in our usage, covers the full range from grammatical knowledge to Gricean rhetorical inference.

The ideal reader may experience Questions at any of the levels of content, genre and text. "Did the butler do it?" exremplifies a content question. One of our texts begins, "Silver was a little pony." At this point the IR has occasion to ask itself whether this is going to be a story of the genre that has anthropomorphized animals as characters or whether the pony is a genuinely animal character in a story about a little boy or girl. (The former turns out to be correct. This exemplifies the eventual resolution of a genre question.) In the same story a textual question is raised by the expression little pony. Is pony to be understood as an undersized breed of horse and little understood as a normal, restrictive modifier, the whole phrase meaning something like "young Shetland"? Or is pony to be taken as "small or young horse (of any breed)" and little to be taken as a nonrestrictive and largely redundant, stylistic modifier? (Cf., little horsie.) This question finds resolution only some time later when the reader is told that "Silver's mother was a big brave horse," and we decide on the second reading.

Closely related to Questions are Hypotheses, which also occur at all three levels. A content hypothesis is, "It was probably either the butler or Major Mud." A genre hypothesis would be exemplified by the postulation after an opening sentence like "Good King Talamain had three beautiful daughters" that the text will probably turn out to be a fairy tale. If such a text continued, "The oldest loved currants. The second one loved crab apples," a textual hypothesis would be exemplified by the IR's postulating that it was about to hear something about the food preferences of

the third daughter. This hypothesis would also exemplify a forward oriented hypothesis or expectation. Hypotheses may also be backward oriented, as in the case of a pronoun whose reference the IR tentatively assigns to one or two or more previously introduced participants.

Questions and hypotheses are subject to strengthening, weakening or full resolution by further developments of the text. We have already considered several instances of questions and hypotheses being resolved. These judgments are also part of the IR's notation.

2.2 Column Two: Properties of the Interpreter

In column two of the three-column document we list the items of the text that serve as justification for the process postulated in column one. To repeat a simple example given above, line 1 of the analysis of Bronco Buster specifies in the left column that a content schema CONTEST is brought to the IR's mind. In the middle column of this line the textual item contest is presented as justification for the entry in the first column. The implicit principle of construal justifying the postulation of the item on the left on the basis of the presence in the text of the item in the center column is that lexical items EVOKE relevant content schemata. Thus the evocation of content schemata by lexical items in the text is one of the implicit properties of the Ideal Reader. Additionally, it is this particular property of the ideal reader that justifies line 1 of our analysis of Bronco Buster.

The first two columns of the three-column document may thus be considered as somewhat analogous to a mathematical proof. The left hand entry represents a "step" in the "proof" and the middle entry represents a justification of this "step." Connecting the justification entry and the process that it justifies is always some principle of text construal. The totality of such principles constitutes the substantive theory of the Ideal Reader. This theory is, needless to say, far from complete; hence we cannot always furnish an explicit principle that derives the postulated process on the left from the justifying piece of text in the center column. Nevertheless, we have the beginnings of such a set of principles, and it is to the discussion of this first-order approximation to the theory of the IR that the present section is devoted. We have found, incidentally, that the exercise of going through a text and attempting to create a document of this sort, in which each processing action of the intuited Ideal Reader must be justified by something explicitly present in the text plus some principle(s) of text construal, is a useful way to get hypotheses about just what the principles of text construal are.

2.2.1 The Parsimony Principle

In discussing the need to keep track of the different levels of certainty of the envisionment, we are implicitly considering

the IR's ability to read between the lines of a text, that is, to draw inferences that are not directly warranted by the grammar and lexicon of the text but that are nonetheless necessary to derive a coherent envisionment. For example, from a text such as (10), the reader who fails to draw inferences such as (11 a-g) will not be an ideal reader.

(10) One day a chef went to Fisherman's Wharf and brought some fish from a fisherman.

(11a) The chef will cook the fish at his restaurant. He will not, for example, take it home to his wife.[9]

(11b) The fisherman caught the fish. For example, he is not selling it for an electrician friend who happened to buy too much fish at the supermarket.

(11c) The fisherman is a commercial fisherman, and he caught the fish in order to sell it. He is not, for example, a lucky sports fisherman who just happened to be opportuned by the chef.

(11d) The money the chef gave the fisherman was not his own money but the restaurant's.

(11e) The purpose of the chef's visit to Fisherman's Wharf was to buy fish, not, for example, to visit the Wax Museum.

(11f) The transaction took place at Fisherman's Wharf. Compare: "One windy day Charlotte went to the hairdresser's and bought a bandana from a street vendor."

Each of the inferences of (11) is of the between-the-lines type, that is, it is not absolutely warranted by the grammar and lexicon of the text. In each case the contrasting possibility mentioned is not what we immediately infer but is something that could turn out to be what the author had in mind. Inferences (11 a-f) are like the middle level inferences (c, d, and perhaps e) in our baseball example (2,3); the ideal reader of this text would draw these inferences but would also remember them as, in varying degrees, subject to suspension by possible later developments. If they are suspended by later developments, the reader will be entitled to a reaction of surprise; the longer the delay, the greater

---

[9] For simplicity we assume here that chefs are employed only in restaurants, though it is of course true that chefs are sometimes employed in the homes of the rich. Another assumption we have made tacitly, that has perhaps been noticed by some readers, but probably not many, is that the chef is a man. Such inferences from stereotypes are a major target of various kinds of language reform.

the surprise.

We must digress briefly from our discussion of inferences (11 a-f) to note that these do not exhaust what the ideal reader should extract from this text. In particular, the ideal reader will construct several little histories for the various participants in this story and know the temporal points of their relation to each other. Thus the money paid for the fish probably originally came from customers in the restaurant who paid for food there and may well end up being paid out for bait or gasoline or fishing supplies. The catching of the fish by the fisherman took place before the commercial event, while the cooking by the chef will take place subsequently, and the eating by the customers of the restaurant later still. Some, but not all, of the relevant historical understanding is represented diagramatically in the following picture (Figure 1).

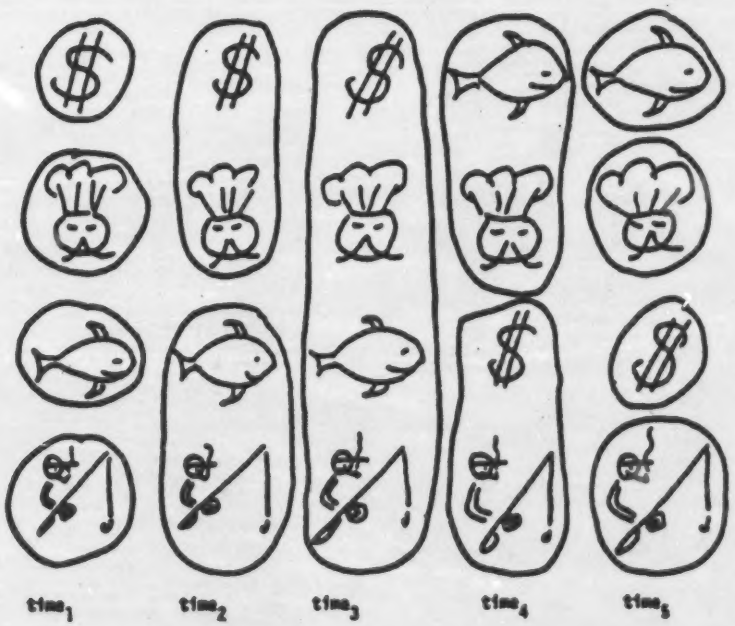


Figure 1

At time-1 the chef has not got hold of this particular money and the fisherman is not yet in possession of the fish. Some time later the money and the chef have become associated and the fish and the fisherman have become associated; this is represented by the picture at time-2. The commercial event takes place at time-3, briefly bringing all four participants into spatio-temporal contiguity. Following this, at time-4, the four participants divide into novel pairings, fish with chef and money with fisherman. Eventually, time-5, these pairings dissolve as the fish is eaten and the money spent or invested.

The relations depicted at times 2, 3, and 4 follow from our knowledge of the commercial event schema itself, while those at earlier and later points in the history, represented by times 1 and 5, depend on our schematic knowledge of the kinds of participants involved in the particular commercial event portrayed in this text. For example, with reference to time-1, it follows from our knowledge of FISHING (not of commercial events) that the fish and the fisherman were not always associated. If we consider a commercial event with different kinds of participants, we are not led to imagine an earlier time when the seller and the goods have not yet become associated. For example, in O. Henry's Gift of the Magi, the sale of the young matron's beautiful hair does not invite us to imagine an earlier time when hair and matron enjoyed separate existences. Similarly, the ideal reader who encounters a commercial event in which someone buys an artificial limb or a pacemaker will not foresee a future time when the purchaser and his acquisition have become separated.

Of the relations depicted at times 1 and 5 in Figure 1, the separation of fish and fisherman at time-1 seems to have the most direct warrant in the text, in the form of a fishing schema which insists that fisherman catch fish they didn't previously have. (An activity, otherwise like fishing that takes place in a bucket or a bathtub is probably not real fishing and surely not prototypical fishing.) In the actual analysis of any text, one always reaches a point at which the inferences that different competent analysts attribute to the ideal reader diverge. For example, some might wish to say that the separateness of the money and the chef at time-1 is something that the reader is supposed to get out of this text, while others will say that this is a plausible imagining, consistent with the text, but not something that would cause us to say that a reader had missed something if he or she failed to come to it. The methodological point is that for any text there are many things that all analysts agree in attributing to the ideal reader and also things about which they will disagree. These latter define those levels of envisionment where what is ordained by the text shades imperceptibly into what is permitted by the text.

In talking of evaluations and resolutions we have considered things that happen to the IR's inferences after they are drawn, but we have not talked about how they are drawn in the first

place. A full theory of the ideal reader will be concerned with the latter quite as much as the former. Although a complete theory of the ideal reader, one which predicts the actual inferences drawn from a text without making use of the human judgment of the analyst, seems distant from the present state of the art, we have been able to take a few steps in this direction. Let us proceed inductively in the analysis of text (10) and the inferences (11) to which it gives rise, first discovering the background schemata that seem necessary and then investigating how these schemata are employed in reaching these inferences.

First, the word chef evokes a RESTAURANT schema. We have already noted that chefs may not always occur in restaurants and we may at times wish to speak of a CHEF schema per se. Hence: in semantic memory schemata may intersect each other in a variety of ways: the CHEF schema is itself an (optional) element of the RESTAURANT schema and is also an instance of an OCCUPATIONAL schema.

There is also a FISHING schema evoked by our text. In fishing there is a person who attempts to catch fish, sometimes succeeds, may own fishing equipment, may ride in a boat, etc. There are two principal kinds of fishermen: commercial fishermen and sports fishermen. One way to give theoretical recognition to this observation is to way that the fishing schema has two subschemata, sport and commercial. A contrasting way is to say that the fishing schema itself instantiates two distinct families of schemata: OCCUPATIONS and AVOCATIONS. A formal theory of semantic memory, one that specifies not only its constituent schemata but also the relations among them, would presumably have to take a position on issues such as this. For present purposes we are content to discover some of the elements of such a theory without specifying their interrelations.

We have made use in a casual way of the concept family of related schemata. In the present case, we may note that for all their many dissimilarities the FISHING and COMMERCIAL EVENT schemata have something in common. They both involve purposeful actions and hence each involves at least one goal.

There is also a schematization of the time relations of actions involved in this, as in any text. (The temporal schemtization of (10) is discussed in Footnote (10).)

We would like now to say something about how these schemata are employed in drawing the between-the-lines sorts of inferences listed in (11). As a preliminary, let us consider how schemata are involved in drawing ground-level, unsuspendable inferences from our text (10). One such unsuspendable inference is:

(12) The person who went to Fisherman's Wharf is the same as the one who bought the fish.

We can put this more long-windedly but in a way that better

reveals our theoretical preconceptions, as follows: the grammar of (10) guarantees that the entity who fills the TRAVELER slot of the JOURNEY schema evoked by the first clause is the same entity as the one that fills the BUYER slot of the COMMERCIAL EVENT schema evoked by the second clause; the TRAVELER participant in the JOURNEY scenario that forms the first part of our envisionment and the BUYER participant in the COMMERCIAL EVENT scenario that forms the second part of our envisionment are the same. We are talking here of schemata as structures in semantic memory that are employed on particular occasions to build the scenarios that constitute an envisionment. We speak of schemata as containing slots and of the things that fill these slots in a particular scenario as participants. Often, the slots in schemata that are filled by participants in scenarios are matched by noun phrases in the grammatical structure and the question whether two participants in different scenarios are the same often corresponds to the question whether two noun phrases are coreferential. We see thus that the main thing happening in inference (12) is that a single participant, the chef, connects two scenarios (going to Fisherman's Wharf, buying a fish) by filling distinct slots (TRAVELER, BUYER) in the schemata (JOURNEY, COMMERCIAL EVENT) that underlie the two scenarios. The two little scenarios are thereby joined into one larger scenario, giving the text coherence.

We have already noticed that, unlike the inferences of (11) with which we are primarily concerned here, the inference of (12) is forced by the grammar. If we consider a text of parallel grammar but which lacks any intuitive semantic support for an inference comparable fo (12), we see that such an inference (14) is forced anyway. That is,

(13) One day a chef went to Fisherman's Wharf and sprained his ankle.

(14) The person who went to Fisherman's Wharf is the one who sprained his ankle.

unlike the inferences (11), the grammar forces the inference of (12) from (10) as it does the inference of (14) from (13).

This aspect of sentence semantics seems directly related to a principle of text semantics that we call the "parsimony principle." (This principle is similar to ideas expressed by Harvey Sacks, Yorick Wilks, and Rick Granger.)

The parsimony principle is this:

> Whenever it is possible to link two separate scenarios into a single larger scenario by imagining them as sharing a common participant, the ideal reader does so.

Let us turn directly to our examples (11) to see how this

principle works out in practice. In (15 a-f) the inferences (11 a-f) are repeated and each is followed by an explanation of how it arises from (10) via the parsimony principle.

(15a) The chef will cook the fish at his restaurant. The food participant in the restaurant scenario and the goods participant in the commercial event scene are the same.

(15b) The fisherman caught the fish. The fish participant in the fishing scenario is the same as the goods participant in the commercial event scenario.

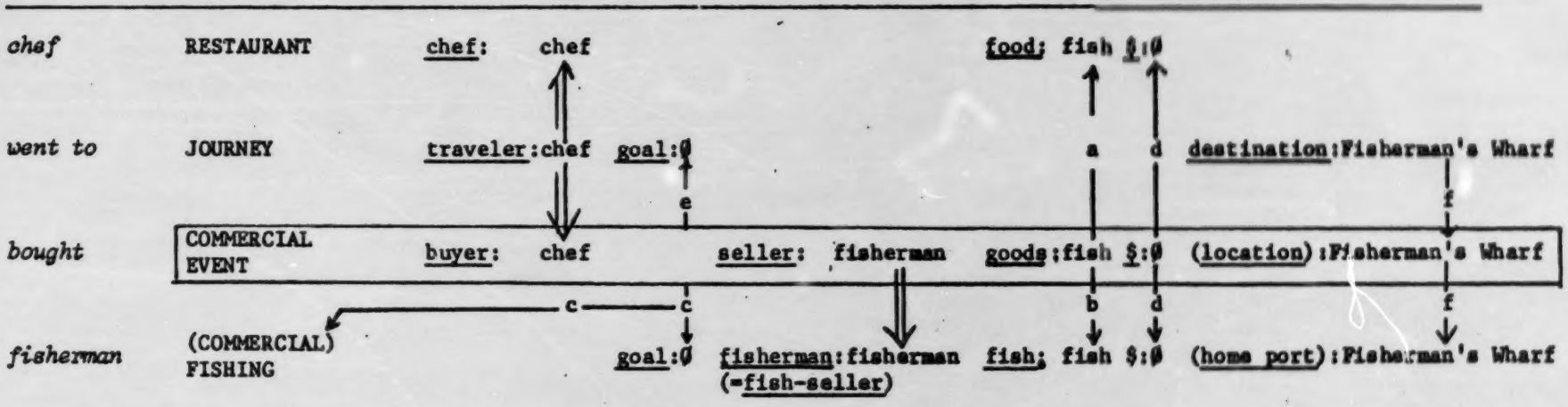
(15c) The fisherman caught the fish in order to sell it and he is a commercial fisherman. The entire commercial event scenario is the goal participant of the fishing scenario. In the commercial fishing schema (or sub-schema) the fisherman is also a fish seller. If the commercial fishing schema is chosen, then the seller slot of the commercial fishing scenario can be occupied by the same participant as the seller slot of the commercial event scenario. (This requires something beyond the parsimony principle as baldly stated; we will come back to this point.)

(15d) The money the chef gave the fisherman was not his own but the restaurant's. The restaurant schema has a slot for money, since restaurants are profit seeking enterprises. The money slot in the restaurant schema can be made to share a participant with the money slot in the commercial event schema.

(15e) The goal of the chef's visit to Fisherman's Wharf was to buy fish. The entire commercial event scenario can serve as the goal participant of the journey scenario.

(15f) The transaction took place at Fisherman's Wharf. The destination participant of the journey scenario is inferred to be the same as the (optional) location slot of the commercial event scenario and the (optional) home port slot of the fishing scenario.

These equations of participants across scenarios are displayed graphically in Figure 2 by the vertical arrows with single shafts. (Only the double-shafted arrows correspond to equations directly mandated by the language of the text and therefore requiring no potentially suspendable inferences.) Words in italics indicate words of the actual text. The words following these in capitals are the names of schemata that are evoked by the words of the text. Following the name of a schema is a series of names of slots in underlined lower case. Following the name of each slot, after a colon, is a word in ordinary print intended to indicate the identity of the participant of that slot in the particular scenario created by this text. For example, the arrow marked 'a' indicates that the <u>goods</u> participant of the COMMERCIAL EVENT



Schemata are evoked by words of the text. Schemata contain slots. Scenarios are instantiations of schemata in which the slots are filled by particular envisioned participants. The grammar and lexicon of the text and, particularly, the parsimony principle, create equations among participants in different scenarios, binding these individual scenarios into the larger scenario that is the envisionment.

Figure 2.

scenario is the same as the food participant of the RESTAURANT scenario.; the fact that this arrow is single shafted indicates that this is a suspendable inference and the fact that it is marked 'a' indicates that this inference corresponds to the inference labelled (11)a and (15a) in the text. The list of slots appearing for a given schema is not intended to represent a comprehensive analysis of that schema but only to portray those slots whose participants are identified with the participants of scenarios based on other schemata in the envisionment of this particular text (by the ideal reader). The inferences indicated by single arrows a - f on the Figure are instances of the parsimony principle in action.

The examples of the parsimony principle so far considered have had the following structure: (1) words of the text evoke schemata, (2) these schemata contain slots that need to be filled by participants in the text-specific scenarios that instantiate them, (3) the parsimony principle enjoins the ideal reader to keep the number of distinct participants to a minimum, hence to equate participants in distinct scenarios whenever possible, (4) this results in the entwining of the various small scenarios, creating out of them one large, reticulate scenario. We would suggest that this reticulation of the overall scenario (i.e., of the envisionment) is an important part of our intuition of textual coherence.

We have concentrated on examples in which the words of the text evoke schemata and the parsimony principle then goes to work on these. But we have also noted in passing certain examples where the parsimony principle does even more: it helps to direct the initial choice of relevant schemata or sub-schemata. One such example involves the choice of the commercial fishing sub-schema over the sport fishing sub-schema. We note that if we choose the commercial fishing sub-schema, then we might get a tighter schematic fit between two participants who must be matched anyway on account of strict grammatical schemata. Recall that the grammar ordains that the fisherman participant of the fishing scenario and the seller participant of the commercial event scenario be the same. By choosing the commercial fishing sub-schema, in which fishermen are fish sellers, we achieve a matching in semantic schematization of participants who are anyway identified with each other by the grammar of the text. Thus, while the fishing schema is EVOKED by the words of the text, the commercial fishing schema is INVOKED by the IR because of its potential contributions to increased connectedness of the overall scenario--i.e., the parsimony. Here we see that the parsimony principle not only matches participants for schemata that are already evoked but leads the ideal reader in his selection of the content schemata themselves. This suggests a sort of metaprinciple for schema selection:

> select schemata in such a way as to give the parsimony principle the widest possible scope of operation.

We might dub this meta-principle the parsimony promotion principle. Much more empirical work is obviously needed to delineate the details of the operation of the closely related parsimony and parsimony promotion principles, but even the very preliminary examples discussed here suggest that these principles are real and play an important role in both ideal readers' and competent real readers' interpretation of texts.

2.2.2 Additional Properties of the Interpreter

In our discussion of the parsimony principle we have had occasion to mention several subsidiary processes.

One of these is the matter of evoked versus invoked schemata. The point of this distinction is that the interpreter of a text is stimulated both by the actual material of the text and by his internal tendencies to find cohension in (or impose cohesion on) the text. So long as the text evokes the schemata necessary to the construction of a coherent envisionment, the reader's action can be relatively passive. To the extent that an author leaves it to the reader to invoke schemata necessary to weld the various elements of the envisionment into a connected whole, we have before us a more demanding text. Thus (16a) is more demanding of the reader as an opening sentence than (16b), even though the reader has to process some more lexical material in (16b).

(16a) John opened his door and saw the road speeding by below.

(16b) John opened his car door and saw the road speeding by below.

In interpreting (16)a, the reader must invoke the idea of a vehicle like a car in order to connect the ideas of John being in something with a door and a road speeding by. In (16)b, the word car directly evokes this schema for the reader.

The contrast between spontaneous and demand computations in AI is closely related to the distinction between evocation and invocation of schemata. When we say that something in the text evokes a schema, we have in mind that the reader upon encountering the textual item is led automatically to call the schema in question into currently active processing space. A demand computation on the other hand is exemplified by the invocation of a schema; the system/reader is confronted with some input that causes it a problem ("How can I visualize looking out of a door and seeing a road going by below?"), it searches in memory for something--a content schema, in the present instance--that will solve the problem. The invoked schema bridges the gap between the otherwise disconnected ideas.

A related ability that IRs have is that of being able to impute cohesion in a backwards direction. In a very easy text, the envisionment is built bit by bit as each segment of the text progresses, with no need for revisions, backtracking, or holding

items in temporary storage while the reader hopefully awaits new information that will enable it to fit them into the envisionment. More demanding texts require revision of their envisionment and may also require of the reader that it hold certain items of information in temporary storage until their role in the envisionment becomes clear. In normal texts of the length of reading test items, these abilities are not usually required as the author necessarily has so little to say that there is unlikely to be a sequencing problem. But reading test items are not infrequently constructed (apparently purposely) so as to tax such abilities in a peculiar way. (From the existence of such peculiar texts, we infer this intention. Since, as we discussed in the introduction, there is no literature relating the structure of reading test items to the abilities they are supposed to test, there is no way to get direct evidence regarding the intentions of the constructors of items. One such test item is the following:

> Pat noticed it at the last possible moment. It was good that she had, because she had almost stepped on it. It was the most beautiful shade of blue she had ever seen. Pat decided it must have fallen out of one of the nests in the tree above. She bent down to look at it and saw that it had not broken. "Oh, good," she thought. "Maybe if I leave it alone, it will still hatch."

In the example just considered, it is a textual item, the pronoun *it*, that has to be held in storage. Sometimes the backwards reading-between-the-lines ability has to do directly with content, as exemplified in the following.

(17) Farmer John planted several apple trees. Fifteen years later his wife's apple pies were the talk of the county.

To make sense of (17) the reader must fill in between the events recorded in the second and first sentences of the text the ideas that the apple trees grew up, produced apples, and that these apples (and no others) were used by Mrs. John to make pies.

Another and very important interpretive ability of the IR's is also related to the parsimony principle, althought not clearly exemplified by examples (11), considered in the last section. This ability concerns the interaction of schemata at different levels. In the story about the little pony named Silver, a moment arrives at which Silver has his hoof caught between some rocks, a hungry wildcat is approaching him, Silver has called for his mother, and we have been informed that his mother is "a big, brave horse." At the level of genre, we are in a narrative with a protagonist (Silver), who has a problem (immobilized and under attack by a predator). Still at the level of genre, we expect that our PROBLEM will have some sort of RESOLUTION, as Resolution is the slot that follows the Problem slot in the narrative genre schema. (This is an oversimplification but not a dangerous one in the present context.) At the level of content we have a new character

introduced, and we are told that she is "big and brave." (Also we infer that she is not in Silver and the wildcat's immediate vicinity, since we have been told that Silver called to her.) The IR at this point contracts the expectation that Silver's mother will come to his aid and chase away the wildcat, thus integrating the new mother character into the main plot of the narrative in a way that will produce a content event (the wildcat's leaving) that satisfies the next slot in the genre schema (resolution). From the fact that the text has only three more lines at this point, we are able to make this hypothesis directly rather than posing the question whether the wildcat will eat Silver's mother thus prolonging or intensifying the protagonist's problem and knowledge of this kind of animal story genre may help the IR reject this possibility as well. But in either case, our expectations with respect to the content (mommies help their fry) reinforce our expectations about the genre (problems get resolved). The ability to integrate information so as to achieve parsimony <u>across</u> levels of analysis, here across the levels of content and genre, is an important construal ability of the IR.

The final interpretive ability we shall consider in this section also involves the interaction of schematizations at different levels, but in this case the interaction is negative and hierarchical rather than parallel and mutually reinforcing. It has been pointed out (e.g., by Wilensky (1982)) that the early "stories" that were considered in the story understanding tradition of artificial intelligence were not really stories in that they had no "point." An example of such a story is the sort of thing we find in (18):

(18) Joe went into a restaurant. The waitress came over. He ordered a hamburger. He paid and left the restaurant.

The principal role of stories like (18) in early theorizing was to demonstrate how an interpreter can use a temporally organized schematization (or "script"--Schank and Abelson (1977)) to fill in missing details of a text. Thus, knowledge of the restaurant script allows both the human and the electronic reader of the text to infer that Joe ate the hamburger.

The usual thing to say about the human and electronic reader's ability to decide that Joe ate his hamburger--as against, say, throwing it on the floor--is that the reader automatically fills in the default values of a schema that has been evoked unless contrary information is presented. This operating principle of interpretive systems has generally been accepted at face value and motivation for it has not normally been sought. But we can readily see what the motivation for this principle is, and taking note of this motivation will do us double service because it will also explain the observation we started out with: that "stories" like (18) aren't real stories because they don't have any point. The motivation in question is that we take this story to be a purposeful communication of a rational being, who shares

with us knowledge of the restaurant script, and so we assume that our author has assumed that we will fill in default values unless he tells us differently. The principle is essentially that of Grice's axiom of quantity: the author doesn't have to tell us that Joe ate the hamburger because he knows that we can figure that out for ourselves. Of course if the envisionment that the author wanted us to achieve had Joe throwing hamburger on the floor, then the author would know that we would need to be told that. This is the obverse side of the same principle that enjoins us when we are told of a man with one blue eye that the second eye is probably not blue.

The moral here is that readers (including ideal readers) do not blindly apply the text, content and genre schemata at their disposal in construing a text but always do so within a framework in which they presume a rational author, conforming to conditions of the general kind elaborated by Grice (or in the case of certain highly sophisticated authors, deliberately subverting or otherwise toying with them). This fact has many consequences for the model of the IR. At any point at which the IR asks himself, "Why am I being told this?" the principle is in operation. We will consider here only one such consequence, one that is especially clearly instantiated in narratives. This is that while at a given temporal point in a narrative the IR is willing to fill in with default values any past script-slot that lacks a value, the IR is most emphatically not willing to do this for future slots. The reason for this is that the IR figures that if everything in this narrative were to go according to the most common default values of the schemata it evoked, there wouldn't be any point in the rational author's telling the story. Text (18) lacks a point precisely because everything we are told perfectly fits the only script it is relevant to. The reader of a narrative will always suspect that expectations produced by a mechanical forward projection of schemata already evoked will be violated sooner or later. Otherwise there is no story, only a completely newsless chronicle of utterly predictable events. One is reminded of the old newspaper wheeze: "'Dog bites man' is not story; 'Man bites dog' is a story." The ideal reader of narratives will expect surprises and the ideal reader of chronicles and expository prose will expect news. The automatic default-filling process works only backwards but not forwards in the processing of a text.

2.3 Column Three: Interview Questions

The questions presented in Column Three of the three-column document represent our best effort at specifying in advance of the interview the questions that can be asked to discover whether or not the real reader being inverviewed had undergone the same processing as that attributed to the IR in the corresponding entry in Column One. These questions were not followed slavishly by the interviewers for a variety of reasons. First, it not infrequently happens that in answering a question posed at one point, the interviewee also volunteers the information scheduled for querying at a later point. In such a case it would be absurd to persevere blindly in asking the questions just as they are listed in the document. For example, the first two questions in Bronco Buster are aimed at getting at the child's knowledge of contests and rodeos respectively. Since the textual stimulus to these questions is the compound noun rodeo contest, many children, in answering the first question about contest naturally used their knowledge of rodeos as a timely and appropriate way to display their understanding of what a contest is. If a child displayed thorough knowledge of what a rodeo was in this way, we did not then, of course, follow up with a question like, "Do you know what a rodeo is?".

A second, and in a way opposite, limitation on the degree to which the questions could be followed verbatim in the actual interviews is the fact that it is impossible to pose a question without presupposing something and when the interviewers had the suspicion that the child lacked the presupposition of the scheduled question, it made sense to alter the course of questioning. The text Strange Machine was difficult for many children and therefore frequently posed this problem to our interviewers. It is reproduced below for the reader's convenience.

> "In 187[illegible], a machine appeared which surprised many people./ Can you guess the name of this strange new machine?/ As you spoke into the mouthpiece and turned the handle,/ a tube covered with a thin piece of tin moved around./ As the tube moved, a needle pressed deep lines into the tin./ As you turned the handle once more,/ the needle pressed against the same lines and played back your words./ This was the first phonograph!/ How different from the hi-fi of today!"

The reader of this text may not be surprised to learn that on reaching the end of this passage many of the third graders were quite confused and had completely lost track of the idea that they were supposed to be guessing the name of the machine from this description. Although we try to avoid as much as possible phrasing questions in a way that presupposes that the interviewee has understood everything perfectly up to that point, it inevitably occurs in actual interviews that our preset questions presuppose information that the interviewee does not appear to have in mind. At such points in the interview, the interviewers has to use his or her best judgment about how to proceed.

In short, the Column Three questions were used as guides by the interviewers to indicate the kinds of information they should find out whether or not the interviewee had. One probe that was used often, at the interviewer's discretion, was this: "Suppose you were watching this in a movie. Tell me in your own words what you would be seeing right now." We found this question particularly useful when the child appeared lost, as a way of finding out just what sort of envisionment had been achieved so far.

The purpose of presenting the material segment by segment and asking questions at the end of each segment was to allow us to trace the envisionment's evolution in the mind of the text's reader. Questions posed after a text has been read in its entirety can, in the best of circumstances, reveal something about the results of the reading process, but such questions can not easily tap into the dynamics of that process. By asking questions after each segment is read, some insight is gained into the temporal development of the envisionment on the part of the reader. Also, such important aspects of the dynamics of the reading process as surprises, expectations, wonderment, resolutions, and so on, are amenable to investigation by a method that segments the reading process and asks questions along the way, rather than waiting until the end before posing questions.

Some of our questions were also aimed at probing certain "meta-cognitive" states (Markman,1979). Is the child aware of lacking this or that piece of information, aware of not having understood a certain word, or bothered because the latest sentence (which, say, the child has misunderstood) does not connect with the envisionment as it has been built up to that point?

The mechanics of the interview itself was as follows. Each child was first exposed to the segment-by-segment reading and questioning procedures with a short text of our own devising, not followed by test questions as in a reading test item. This got the child used to our basic method of presentation. Then the child was exposed to two reading test items, one in the strict segment-by-segment presentation form and the other in the "silent read-through" format--that is, where the child was first asked to read through the item silently, and then was asked to read it aloud, segment by segment, and respond to the interviewer's questions after each segment. (As mentioned in the introduction to this report, in the "silent read through" procedure, the form of certain questions was necessarily altered: e.g., from "What do you think will happen next?" to "When you read this before, what did you think was going to happen next?".) Normally the interview with a given child was performed in two sessions of 30-45 minutes each, with the break coming between the two reading-test items. Half the children were given the "silent read-through" treatment first and the other half the regular procedure: no effect of this order on the children's performance was observed.

3.0 Results

We will discuss first, in section 3.1, our results with respect to the text materials and then, in section 3.2, our findings from the interviews with the children.

3.1 Critique of Test Materials

This section will be organized into three subsections. Section 3.1.1 will present a critique of selected questions. Section 3.1.2 is a summary of the remarks that have been scattered throughout this report constituting a critique of the texts presented in reading test items. Section 3.1.3 is a discussion of the functioning of reading test items within the pragmatic situation "taking a reading test."

3.1.1 Critique of Selected Questions.

We will present here certain types of questions that we consider in one way or another unfortunate. We will give our reason for the negative judgment. Usually only one example per type is offered, occasionally more, but the reader familiar with the kind of material under discussion should be able to supply additional examples from his or her own experience. The texts referred to in this section are presented in Appendix A. Each question offered as an example is presented at the time it is discussed. Often the discussion of a given question will be extended to cover more than that question's most obvious fault.

1. Questions should not test for abilities not directly related to reading.

Example: "Strange Machine." Question 1.

About how many years ago did the machine first appear?

(1) 25 (2) 50 (3) 100 (4) 200.

To answer this question the child must know the year in which the test is being administered, must subtract 1877 (mentioned in the text) from that figure, subtract the absolute value of that difference from the absolute value of each of the four candidate answers, find the smallest of the resulting differences, and choose as answer the candidate that produced this smallest difference. The skills involved in performing these operations are not properly considered reading skills. This question is gratuitous for two other reasons: (1) the length of time since the first phonograph appeared is not a detail that is directly relevant to the theme of the passage, which concerns the extent to which the physical appearance and operating characteristics of the original machine differ from its modern descendants. When a question to a reading comprehension passage seeks to determine the degree of detail that the reader recalls (or can find again by looking back), details that are part of the central theme of the passage

should be questioned, not irrelevant details. The skill touching on recall of detail that we hope to impart to the fledgling reader is to find and retain those details that go to make up the main theme or point, since nobody can retain all details of what he or she reads. (2) The length of time passed in years is not even remotely relevant to the theme of the passage although the time gap in terms of the changing socio-economic and technological environment is. There is no reason, however, to believe that a third grader can connect the phrase "about a hundred years ago" with any sort of picture of the style of life of that time, nor is it reasonable to categorize this inability as a deficit in reading ability.

2. Reading comprehension test items should not test for knowledge of vocabulary per se but rather for more integrative skills. Vocabulary can be and is easily testable separately.

Example: "Strange Machine." Question 2.

2. How did people feel when they saw this new machine?

(1) angry (2) astonished
(3) worried (4) frightened

The desired answer is astonished, a word it is unlikely that all third graders will know. Since the hard word occurs among the candidate answers, rather than in the text, a child could understand the text perfectly and yet get the question wrong, a problem to which we will return below. Moreover, this particular question is one among many which penalizes the child who has some sensitivity for language--the last thing one would wish a reading test to do. Astonishment is an extreme degree of surprise, which latter is the more general term. Hence it is possible to be surprised without being astonished, as it is possible to be hungry without being ravenous. A child sensitive to these facts of English might on that account interpret the candidate astonished as a clever trap and choose instead one of the other possibilities, one that is at least not contraindicated by anything in the text. Test wisdom tells us of course that this is not the kind of reasoning testers imagine or reward, but test wisdom is not reading ability. This question, like the previous one, also illustrates the fault of testing for a detail that is not part of the main theme or point of the text. The test makes nothing further out of people's emotional reaction to the new machine mentioned in the first sentence.

3. Questions should not employ words that are unfamiliar to children at the relevant grade level or that are unfamiliar to them in the intended sense.

Example: "Nasrudin." Question 3.

3. Nasrudin suggested there were no trees in the Sahara Desert because--

(1) trees can't grow in the desert
(2) no one had ever planted any there
(3) they had been destroyed by fire
(4) he had chopped them all down

This use of the word suggested, meaning roughly "implied" or "mildly asserted," is unlikely to be familiar to third graders, though of course suggest in the sense of "give advice" will probably be familiar to them.

Example: "Bronco Buster" Question 1.

1. A bronco buster who ignores a rodeo rule is

(1) skillful (2) disqualified
(3) chosen (4) a winner

The word ignore is unlikely to be known by many third graders. (In fact we found in our interviews that it is not known by many third graders. We do not yet have extensive interviews on the Nasrudin story cited just above.) Once again, a child could understand the passage perfectly and get the question wrong because he didn't know the word ignore that occurs in the question part of the item. In general we would offer the following maxim to constructors of reading tests:

> Construct questions (and answer candidates) so that if the child understands the passage he or she will get the answers right and if the child does not understand the passage he or she will (probably) get the answer wrong.

In particular if the question is harder to read than the passage, it will often occur that a child understands the passage and gets the item wrong. (This applies not only to the lexical items occurring in the question.)

It may be objected that one might just as well consider the PASSAGE + QUESTION unit as constituting the test item. This, however, will not do because it offers no way of distinguishing reading ability from test-taking ability. In the limiting cases, this formula will admit as a perfect reader the child who understands nothing of the passage but through test-taking acumen gets all the answers right, and as a non-reader the child who understands the passage perfectly but, because harder material occurs in the questions, gets all the answers wrong. These reductio cases are unlikely to occur in actuality. Still they point out the utter insufficiency of the idea that the PASSAGE + QUESTION unit is a theoretically reasonable candidate for the unit of testing. The goal in the creation of reading comprehension tests must be to ensure that understanding of the passage is an accurate predictor of performance on the test. Only then will we be in a position to

select the passages that appear on reading tests as representative of the kinds of reading materials we expect a child of the given age to read with understanding.

In this discussion of unfairly difficult questions we have considered a kind of fault in item construction analogous to statistical errors of Type II: Our test may lead us to conclude erroneously that the world is not the way we want it to be (i.e., that the child cannot read the material). The type of testing error corresponding to the statistical Type I error is also known to occur in tests of reading comprehension and has been studied under the rubric "passage dependency." For example, if the answer set consists of the choices irritated, annoyed, pleased, bothered, the test-wise reader (as we discovered in a brief informal pilot test) will be able to choose the correct answer without having recourse to the passage. (In this case, without even having recourse to the question!)

4. A question should have at least one correct answer.

Example: "Strange Machine" Question 2.

How did people feel when they saw this new machine?

(1) angry (2) astonished
(3) worried (4) frightened

We have already discussed the fact that the intended answer astonished is not a correct one. It is preferable to the other answers only on the basis of rather sophisticated and hard to specify assumptions about the intentions of testers. The child must be able to reason somewhat as follows:

> "Astonished strikes me right off as an attractive possibility because it is similar in meaning to surprised, which appears in an appropriate place in the text. But astonished is probably not the right answer, rather a clever distractor, because astonishment is an extreme degree of surprise and if they had MEANT astonished they would have SAID it. That is, if astonished was what they meant, what they said was some kind of understatement, as if they were being sarcastic or something and there are no other cues here to support a hypothesis of sarcasm or irony."

(The fancier, theoretical restatement of this part of the child's reasoning is that, according to Gricean principles, if the author is being cooperative and literal, he will not use a word indicating a less marked value of a semantic variable if he intends to communicate a more marked value of that variable because he knows the addressee will always choose the less marked over the more marked interpretation, given the choice.)

> "So I guess I'll have to look again at the other candidates. Although none of them is directly warranted by the text, each

is possible on the basis of what I know, at least from that sentence alone. A new machine could make people angry, or worried, or frightened, for example if it were a weapon. Yeah, but this wasn't a weapon, it was just something that turned out to be a record player. Still I don't know how people in 1877 would react to a record player. I think they were very religious back them. Maybe they were angry because it would be against their religion because it could play dance music. ... Hey, I've been thinking about this too long, and considering too many hypothetical things that aren't directly in the text I've read. Although astonished is wrong in one way, it is very like a word in the text and I seem to have to go too far afield to connect a less repulsive answer to the text. I'd better pick astonished and move on."

The problem with requiring this judgment of the test taker, regarding the tenuousness with which a plausible answer is connected to the text, aside from its monumental inexplicitness and the fact that it has nothing to do with reading ability per se, is that finding the CORRECT answers to many questions requires the application of highly text-independent pieces of background knowledge. Some of these cases will be discussed under a separate heading below. At the moment we may just point out that while the test taker is at some times required to suppress background knowledge, he is at other times required to activate background knowledge, in order to get the "correct" answer. The rules for when to suppress and when to activate background knowledge on reading tests do not appear to be derivable from the principles governing the applications of background knowledge in normal reading, if indeed there are such rules to be discovered.

Example: "Strange Machine" Question 6

6. Which of the following does this story tell about the machine?

(1) how it worked
(2) who invented it
(3) where to buy it
(4) how many people saw it

The desired answer is how it worked. The passage does not, however, tell the reader how the machine worked. What the passage does is provide a description such that if a reader already had a clear idea about how the machine worked he or she might be able to recognize this as an incompetent attempt to express that knowledge. Again, to get the "right" answer the test taker will have to engage in complex reasoning, an essential part of which is the recognition that this is a reading test, not real reading. This question is not quite as bad as the one discussed just previously because the question explicitly asks about what this story tells us, thus rendering plausible distractors less justifiable. Nevertheless, there is no correct answer, and one must take the question, not literally, but according to the peculiar and never

explicit pragmatic conventions governing the reading test situation in order to get the answer right.

5. A question should have at most one correct answer. (Taken with suggestion 4, this amounts to the familiar idea that a test question should have exactly one correct answer.)

Example: "Dear Dad" Question 2

2. How does Tam probably feel about the trip?

(1) sad (2) worried
(3) proud (4) excited

The desired answer is (4) excited, and were it not for the presence of the distractor proud, this would be a good question, because it requires the reader to draw from the expression can't wait and the exclamation point at the end of the last sentence the kind of inference about the emotional state of a personal letter that a competent reader draws. There is, however, another very likely inference from the passage: that the writer of this letter is a child (traveling with her mother to her grandmother's house and writing to her dad). It would be normal for the nine-year-old reader to think of the writer of this letter as also a child of about that age. The achievement of participating in fishing with her mother and then herself cooking the resulting catch would be an occasion for justified pride for a nine-year-old. The test requires the reader to find more salient the inference of excitement connected to the whole trip than the equally plausible inference of pride connected to its centerpiece event. This is, to be sure, the salience ordering that we as adults impute to these two inferences, but it is not at all clear that the reason we impute the relative saliences in this way is because we are better readers. It is easy to imagine a nine-year-old reader of this text who draws clearly both the inferences of pride in the fish-cooking and excitement about the whole trip, but who empathizes most strongly with the more concrete emotion of pride deriving from the more concrete experience. A reader who is affected this way by a text and is then questioned about the emotional content of the text is within her rights to answer in terms of the emotion with which she felt the strongest connection. Selecting the stipulated correct answer requires the reader not only to understand the story as a competent adult would, but also to take an attitude toward the events of the story that an adult would take. This does not seem a fair test of the child's reading ability.

Example. "Strange Machine" Question 7

What is the story mainly about?

(1) guessing a name
(2) speaking clearly
(3) the first machine
(4) the first phonograph

Some members of our research group who were not familiar with reading tests at the time thought the stipulated answer to this question would be guessing a name. An interesting fact is that the members who were familiar with reading tests immediately and laughingly told the first group that this could not possibly be the answer the test maker had in mind as correct, and of course they turned out to be right: the "correct" answer is (4) the first phonograph. The members of our group who erroneously picked guessing a name thought the testers were probing for the child's apprehension of the structure of the text. It is set up as a kind of riddle in which the general category of a thing is given (machine), the addressee is asked to guess the specific kind, and then (allegedly helpful) facts are serially presented until finally the answer to the puzzle is revealed. Admittedly, "guessing a name" is not the most felicitous choice of words to describe this kind of text structure, but we had already learned that successful takers of reading tests will not be helped by an assumption that test makers use the English language perspicuously. The likelihood of the answer guessing a name, we figured, was enhanced by the unhelpfulness of the "description" offered in the text. It would seem that a child who took the passage seriously and was continually engaged in a, no doubt unsuccessful, effort to guess the name of the machine while reading it would be justified in concluding that the point of all this was the playing of the guessing game, especially as there was no real description of the machine. Moreover, we thought that the intent of the question was to see if the child could reflect consciously on this reading experience. But we were wrong, presumably due to naivete about reading tests rather than naivete about reading. Once again, the choice of the desired answer depends not on the ability to read but on knowledge of some implicit conventions covering the selection of "best" answers on reading tests.

6. Questions on reading tests should not require the application of specific items of world knowledge that are neither given in the text nor of a sort that one may safely assume any normal child of the given grade level will possess.

Example: "Bronco Buster" Question 2

2. In a rodeo contest a bronco buster must keep one hand

(1) under (2) still
(3) free (4) hold

To realize that free is the correct answer, it is necessary to know, aside from anything said in this text, what the essential nature and point of a bronc riding contest is. The hand that the text describes as "in the air" is of course not held in the air for aesthetic effect. Rather the rule is that only one hand may be used to hold on, and it is the hand that is not used for holding on that the text refers to. The reason for this rule, which also needs to have been learned from the reader's text-external experience, is that the bronco is trying to throw the rider off

and the rider is trying to stay on nevertheless. The young reader who knows little or nothing of bronc riding contests, of which we found a considerable number, is likely to envision the scene as more like a dance than a fight, that is, as something resembling a circus riding act or a horse show performance--in which horse and rider are engaged in a graceful, cooperative enterprise. In such a scene, the requirement that one hand be held aloft would probably be motivated by aesthetic considerations, and the hand would quite possibly have to be held still. In any case, the idea of holding one hand free is unmotivated in a scene which lacks the necessity to hang on that derives from the competition between horse and rider of the bronc riding scenario. From what is said in the text, if one has no idea of a bronc riding contest and is unfamiliar with the word bronco buster, the cooperative horse and rider scene is quite as reasonable an envisionment as the competitive horse and rider scene. Thus the route to the correct answer free can only be traced through world knowledge of rodeos in general and bronc riding contests in particular. This kind of knowledge is not properly considered part of a child's reading ability.

It is important to draw a distinction here between a reader's having knowledge of the relevant content schemata and his being able to activate this knowledge in the appropriate way when processing a text. If the reader fails because she lacks the background knowledge, this is not a failure of reading ability, but a reader who has the requisite background knowledge and cannot, nevertheles, harness it to the task of construal in the appropriate way, does display a deficit in reading ability. For example, a young reader who was familiar with the bronc riding contest scene might get this question wrong because of a genuine inability to read the passage, for example not being able to identify the rider of the second sentence with the bronco buster mentioned in the first sentence and in the question.

Because one of the main abilities of competent readers is precisely that of being able to bring world knowledge to bear in an appropriate way, it is neither desirable nor even possible to construct passages for tests that require no world knowledge for their correct construal. Nevertheless, as we have just pointed out, it is essential if the results of the tests are to mean anything, that the children who take the tests be possessed of the world knowledge required for construal of the passages. If the test is to test reading, we must be able to be sure that failure to bring a piece of world knowledge to bear in the required way reflects a processing deficit, not simply the absence of the content schema from the child's knowledge of the world.

In the last decade or so, the testing industry has taken account of this problem in a few selected content areas, namely those in which knowledge of the world is judged to be unevenly distributed across officially identified groupings of people such as ethnic groups and sexes. But it has not been recognized in the industry--or if recognized not acted upon--that the same

principle--viz. a child should not be penalized in a reading test for lacking knowledge of the world--should be applied to all kinds of world knowledge, not only that differentially distributed across politically identified social groups. If Jane reads Dickens while Johnnie watches the rodeo on TV, we don't want our tests of reading comprehension to penalize Jane for the way she spends her time.

7. Questions on tests of reading comprehension should not require the taker to reject distractors on the basis of his or her intuitions about grammaticality.

The example just considered will serve to illustrate this point. Consider the answer hold. This could be a correct answer if the sentence *In a rodeo contest a bronco buster must keep one hand hold were English and meant something like, "In a rodeo contest a bronco buster must keep hold with one hand." If the sentence were English, the gloss offered is probably about what it would mean, so the issue comes down to the judgment by the young reader that this sentence is not English. The fact that this sentence is not English may not be nearly so obvious to a child as it is to us, and this lack of obviousness need have nothing to do with the ability to read. English, together with most languages, is to a much greater extent than commonly realized, composed of a large number of idiomatic and semi-idiomatic expressions governed by highly context-specific and exception-ridden grammatical rules. We may note, for example, that the normal English phrases make one hand bleed or see one hand move have the same morpho-syntactic structure as the non-English phrase keep one hand hold. Also, there are expressions such as put on hold, keep one hand up, and keep one hand moving, each of which is English and each of which contains most of the grammatical elements of the proscribed *keep one hand hold. If the reader of this report will try to specific exactly how English lets these expressions in but keeps the bad one out, he or she will find it an interesting exercise.

Against this background realization of the huge number of pieces of ad hoc knowledge that lie behind judgments of grammatical acceptability, consider now that one of the child's principal experiences in school is to undergo a continual refinement of what he considers grammatical English. On the one hand he finds that certain of the rules he has learned in casual speech are overgeneralized or that they have unpredictable lexical exceptions, e.g., that when you do to a ham what is called roasting if you do it to any other kind of meat, it is called baking, or that while pinkish and purplish are words, *orangeish isn't. On the other hand, she learns that there are many new and special ways to say things that she didn't know, for example that carrying coals to Newcastle is a way of describing a particular kind of unproductive activity or that certain grammatical processes can be combined iteratively to produce such hard to process but undisputably English sentences as David was believed by Ronnie to be suspected by Nancy of being in the pay of Communist dress designers.

In short, the grammar school child is continually learning that what he or she may have thought was the full range of acceptable expressions in his language bears emendation in both positive and negative directions. (Maybe this is why it is called 'grammar school.') The child in school is constantly learning that what he or she thought about English--especially written English--is in some new way wrong. It is therefore unfair and unrealistic to expect a child when offered a string of words like *to keep one hand hold to reject it as non-English rather than, say, to judge it an acceptable but unfamiliar idiom. *Keep one hand hold meaning "keep hold with one hand" could easily be acceptable English. The third grade child is in no position to know that it isn't just one of the many ways of using English he or she runs across every day in school. (The reader who understands the bronc riding passage very well knows that if one hand has to be held in the air, the rider must indeed hang on tightly with the other hand. The rider can't hold both hands free; he must keep one hand "hold." The sense of "must" intended by the test writer repeats that of the rule mentioned in the passage. The sense perceived in "must keep one hand hold (in order not to fall off)", however, is the result of a fairly deep understanding of the point of the passage.)

The experience of being reminded daily that one's own linguistic practices are not adequate for the school context, common to all school children, must be especially exacerbated for bilingual children and for those who speak minority and (other) stigmatized dialects. For these children school is a place where above all else one learns that whatever is one's natural way of saying something is probably "wrong"! How is a young speaker who is under constant pressure to exchange his or her vernacular speech in favor of the "better" form of English that is taught in school going to know whether *keep one hand hold is or is not a part of the prestige dialect that he is constantly being reminded he hasn't mastered?

8. "What is the best title?" sorts of questions should not be asked.

Such questions are almost perfectly ambiguous between reading tests and real life interpretations because the reading test desideratum for an answer is always for the most descriptive summary of the text's contents while in real life (and literature) this is an anti-desideratum for a title except in quite restricted genres such as articles in learned journals. For example, would anyone consider Death, Indecision and the Oedipus Complex in Ancient Denmark or A One-Legged Whaler's Self-Destruction Through Obsessive Pride and Vengefulness to be better titles than Hamlet and Moby Dick?

Example: "Fritz" Question 5

5. Which of these is the best title for this story?

(1) Fritz Takes a Walk.
(2) Fritz Asks a Question
(3) Fritz and His Neighbors
(4) Fritz and His Hats

Competent adult readers not only have a hard time selecting the best answer to this question; they even find it hard to guess which answer the testers had in mind. When told the right answer, Fritz and His Hats, most adults can be made to accept that at a very concrete level this answer does best summarize the content of the story. At a less concrete level, however, the importance of the hats is the effect they have on the neighbors and Fritz's social position in the neighborhood is the larger theme of the story. Consequently, there are some who insist that the "Neighbors" choice is the best. But since good titles to real stories don't summarize the story's contents anyway, none of this is relevant to the question taken literally. Thus, we would expect the child with some literary sensitivities to eliminate both of these summarizing candidates. We have here an example of a question for which, given perfect comprehension, test wisdom is a necessary but not sufficient condition for getting the right answer.

9. Questions should not require as correct answers false propositions that are stated or implied by the passage but not marked in the passage as false or at least surprising.

Example: "Baseball" Question 4

4. Which of the following best describes baseball attendance?

(1) greater than basketball
(2) not increasing like ice hockey
(3) increasing about as fast as that of football
(4) second only to football

The stipulated answer is (2), which may have been the best answer when the item was written, but (1) would be much more clearly true today since baseball attendance has been increasing markedly for the last several years while the current decline in basketball attendance is a matter of current comment. Care should be taken either to make sure that the factual claims of test passages are true or to phrase the questions in a way that makes clear that what is being questioned is that part of the reader's knowledge that has changed (or been challenged) since reading the passage. Test-wise children should, of course, take as an implicit preamble to each reading test question something like "How would a person who believed what I have just read answer the following questions?". But the wit to construe the test-taking task in just this way is not necessarily a proper part of the ability to read.

When we recall that one out of thirty-odd chilren knew the word phonograph and none the word hi-fi (usually pronounced by them "hee fee") in "Strange Machine"--though they all knew the

words record player and stereo- one suspects that not enough attention is paid to the timeliness of the items.

In this critique of reading test questions, we have concentrated on the faults we have found in the questions that appear on the tests. An equally important way to approach the subject, though for obvious reasons a more difficult one, is to consider the questions that should be asked but are not. On this subject we have considerably less to say, but a few words may be in order.

We have two main criticisms regarding the kinds of questions that are lacking. First, there are not enough questions that get at the main point of the passage. (Best title questions probably have this objective, but they fail to achieve it, for reasons already discussed.) For example, both "Nasrudin" and "Fritz" are jokes--of a sort. To realize this, the reader does not actually have to be amused; but the competent reader would recognize that these texts are intended as jokes. To realize that a text is intended to be a joke is one thing, to know what the point of the joke is supposed to be is another, and to find the joke funny is yet a third. None of the questions posed to either "Nasrudin" or "Fritz" deal with any of these three things. Some even seem to fly in the face of the essence of the genre, as in Fritz's Question 6, "What will Fritz probably do next?", next here meaning after the events in the story are over. It's in the essence of a joke that it's over when it's over. A question that might be asked about "Fritz" could be:

Someone might tell this story if they wanted to

(1) make other people laugh
(2) help someone get used to a new neighborhood
(3) encourage people to buy hats
(4) teach people not to wear too many hats

It may be objected that this question would be unfair on the grounds that, although "Fritz" is obviously a member of the genre "Joke,", the story is not one any real person would try to get a laugh with. Perhaps the remedy to this sort of objection is, not to reject questions that seek the reader's appreciation of the genre the text represents, but rather to require that texts represent good examples of their genres.

3.1.2 Critique of the Reading Test Passages

There have been many occasions in this report to mention ways in which passages taken from standardized reading tests failed as suitable representatives of English writing, given the purposes which they were to serve. In many cases, the texts have characteristics which, we felt, unfairly predispose their readers to reach an "incorrect" understanding of them.

Literary texts of several types, as well as overheard snatches of conversation, frequently contain "referring expressions" whose antecedents are not apparent to the reader or the eavesdropper. In narratives that are designed to represent a "point of view" the effect of such a device is to induce the reader to hold certain things in abeyance, confident that if knowledge of the reference is important, it will eventually get revealed. In riddles, of course, the reader, or listener, is required to guess. While the ability to deal with such genres is an important ability, it may be the case that reading test items concerned mainly with the reader's willingness and ability to postpone identification of referring expressions are not good tests, in the early years, of the ability to read.

One of our passages received, in our group, the name "Blue Egg." It goes as follows:

> Pat noticed it at the last possible moment. It was good that she had, because she had almost stepped on it. It was the most beautiful shade of blue she had ever seen. Pat decided that it must have fallen out of one of the nests in the tree above. She bent down to look at it and saw that it had not broken. "Oh, good," she thought. "Maybe if I leave it alone, it will still hatch."

By the middle of the passage the talented riddle-guesser will have decided that "it" is either a baby bird or an egg. The word broken might, to readers sensitive to standard word collocations, rule against its identity as a bird; but it is only the very last word, hatch, which ultimately settles the question. Yet, as we shall see, if the child is confused at the experience of reading such a long passage without knowing what is being talked about, the word hatch might still lead to the wrong answer.

This passage is not necessarily a bad item for a reading test; but we feel that the nature of the child's engagement with the passage would be more like a "real" reading experience if the passage were explicitly identified as a riddle. The child who fails to construe the task as one of solving a riddle, and who is therefore uneasy about not knowing what's going on, is being unfairly treated.

A passage our group referred to as "Amelia" is one, among many others, which had personal pronouns introduced out-of-context, giving the reader, once again, the experience of knowing

that something is being referred to which one cannot yet identify. The passage has several "mid-text" features in its first sentence, which is:

> As she rolled to a stop in the center of the business district, she was quickly surrounded by townspeople.

The use of the unintroduced simple past, the pronoun she, and the definite noun-phrase the center of the business district, all, in ordinary language, presuppose prior identification of anchoring reference points. In this case (as we shall see below), it is not the identity of the "she" which is at issue, but the envisionment of the whole scene. This passage, too, is a kind of riddle.

In a number of the passages we worked with, the final appropriate envisionment depended on the correct understanding of one word. In "Blue Egg," that one word was hatch. Suppose that a child has been reading through this passage and has become confused at the kind of guessing game which it creates. The inability to make any secure identifications makes it impossible to create any sense of coherence in the text, so that at the end the reader is aware only of the need to decide whether the thing is a bird or an egg. Let us assume that by the time the child finds the word hatch he recalls that birds hatch, and chooses to identify "it" as a bird. This decision would come as the result of confusion about the task, because if the child had noticed that the collocation of bird with hatch doesn't work in the future, the mistake wouldn't have been made. ("The bird hatched" is okay, but "This bird will soon hatch," with "this bird" an independently identifiable object, is strange.)

In the way that in "Blue Egg" the final decision could only be made on encountering the word hatch, in "Amelia" the word cockpit seems to be critical. The full Amelia text is as follows:

> As she rolled to a stop in the center of the business district, she was quickly surrounded by townspeople. They were surprised when a woman stood up in the cockpit. When she raised her goggles they were even more surprised. She had been flying into the sun all day and her face was sunburned a bright red except for the white circles around her eyes where her goggles had protected her. She blinked out at the assembled population of Hobbs like a boiled owl.

This is clearly a text that requires some very imaginative invocation of an appropriate situating schema, just in case the word cockpit (which many of our students did not know) fails in achieving the proper schema evocation. The word fly is used in the passage, but late enough, it would seem, to have little effect on the initial schematization (of, say, a car or a motorcycle) which the children would have settled on.

We in no way wish to challenge the test-writers' right to provide items which require the kind of cooperative "filling in"

that an item like this demands; but we feel that there are too many other things wrong with this text to let so much depend on the readers' knowledge of a single lexical item.

The "Strange Machine" passage, discussed frequently above, could be given sense only when the reader came, very late in the text, to the word phonograph. As has been pointed out already, almost all of the children in our study did not know this word, and were thus in no way able to construct an appropriate envisionment of the text.

A number of our texts created envisionment conflicts of one sort or another. One such conflict can be seen in the "Shipwreck" passage, which reads as follows:

> The boat was pounding on the rocks. The passengers were in an uproar. Women were screaming and men were yelling at the top of their voices. The captain of the boat went quietly from one place to another lowering the boats. If you looked at him you could never tell that the boat was about to sink.

There are both auditory and visual incoherencies in this text. We are told that the boat was "pounding on the rocks," suggesting a violent storm. The passengers were hysterical, screaming, milling about. In such a context, one wonders, what could it mean for the captain to move "quietly from place to place"? One can move quietly only in a quiet environment. A number of the young readers of this passage misread quietly as the more coherent quickly. But this, of course, made it difficult for them to decide that the captain was calm when it came time to answer the questions.

The visual incoherency comes from the difference between the distant view of the scene that one creates while reading the first few sentences (we "see" the boats, the passengers as a mob, etc.), and the close-up view of the hero "calmly" performing his captainly duties. The last sentence, "If you looked at him you could never tell that the boat was about to sink," requires us to retain the close-up view; in fact, if you looked at him, you would see that he was lowering life-boats into the sea in the service of his hysterical passengers. The sudden change of point of view, following upon the tendency to misread quietly as quickly just discussed, makes this transition basically incoherent.

In the Amelia text we saw the pilot looking like a boiled owl. The image the author had in mind was apparently a blend of two things: from the fact that the pilot was badly sunburned, we might get the idea that it looked as if she was boiled; and from the fact that her goggles left unburned rings around her eyes, we have the idea that she resembled in some way, about the eyes, an owl. But of course we cannot believe that, when one puts these two images together, she actually looked the way an owl might look which got boiled.

In many cases our texts were critically "uncooperative." In the text we called "Banana Trees" we are given some surprising information, and then nothing is done with it. The text is as follows:

> Did you know that bananas don't grow on trees? Even though the banana plant looks like a tree, it has no trunks or limbs. Bananas are grown in South and Central America and India where the climate is hot and damp. etc.

Somehow, the reader is expected to form the image of a banana tree from the information that it "looks like a tree," and then one is expected to modify that image in such a way as to make it without trunks or limbs. That is not easy to do, so one has the right to expect the text to tell us more.

Some texts require the reader to make rapid adjustments in the details of the envisionment, adjustments that keep one busier than one ought to be. This phenomenon was particularly striking in the passage we called "Camping", reproduced here:

> A strange noise wakes me up. I am very quiet, listening. I hear the wind blowing in the branches above the tent, and I can also hear footsteps. Some paper is rattled. All of a sudden a garbage can lid is tossed on the ground. I jump from my bed. My flashlight reflects two pairs of bears' eyes in the light beam, one from each trash can. This is so exciting to a city boy that I find it hard to go back to sleep in my sleeping bag.

At first, the normal reader puts the narrator in an ordinary bed, in a house. (Level E-1 envisionment.) Then this reader learns that the narrator is in a tent, and is probably inclined to make it a small tent. Then we learn that the tent is a large one, since the narrator jumps from his bed. (Bed, not cot.) But in the end the narrator returns to his sleeping bag! The image we have at the end is one of a large tent, a bed in the tent, and a sleeping bag on top of the bed. To the extent that the reader is attentive to all of these changes, the experience is distracting in the extreme.

A large and important incoherency in the "Amelia" passage was the sudden shift in point of view between the first two sentences. In the first sentence we have the simple definite pronoun she, suggesting that we are experiencing this episode from her point of view. In the next sentence, we are suddenly brought to the point of view of the townspeople, who were surprised at her appearance. The sentence "They were surprised when a woman stood up in the cockpit" is supposed to convey the idea that, while they knew that somebody was piloting the airplane, they were surprised to find that that somebody was a woman. But we are not prepared for such a sudden shift of point of view, since we knew from the start that it was a woman. None of the children in our study who saw this text understood the point of that second sentence; nor did many of

its adult readers. It is a poorly constructed paragraph.

Lastly, we might point out that we were unhappy, in a number of our passages, with certain lexical choices. The oddity of chopping lumber as a description of cutting down trees to make lumber has already been pointed out, as has the unfittingness of the term bronco buster in a context in which bronc rider would have been more approprate. To these complaints might be added the use of the word machine in the guessing game about the phonograph. One of our investigators took it on himself to refer to the household record player as a "machine" for a few weeks after we encountered this passage, and found that his children rejected the word. (It was not tolerated in such expressions as, "Should we put the machine over here?" or "Let's turn the machine off now.")

3.1.3 Pragmatics of the Reading Test Situation

In previous sections we have noted various ways in which the young reader confronted with a reading comprehension test finds himself in a situation different from any usual situation that calls for reading. First, with respect to the association of genre with purpose (e.g., jokes with entertainment, essays with instruction, etc.), we have in the reading test a genre in which the purpose is unclear. To be sure, one may confidently state that the purpose of reading the text is to be able to answer the questions correctly. But as we say in section 3.1.1, the kinds of questions encountered and the relations of these questions to the texts they ostensibly query are so scattered and capricious that knowing one is to be faced with "the kind of questions one finds on a reading test" is to be provided with fundamentally vacuous guidance for reading the text. In the rare case where the question and the passage are carefully articulated, this appears sometimes to result from the tester's desire to punish the child who employs strategies characteristic of good readers. We discussed one such case above (section 1.2) in somewhat different terms: the punch line to the Nasrudin text comes in the final sentence, in which Nasrudin indirectly justifies his claim to be such a great wood cutter that he created the Sahara Desert out of a former forest.

> "That's ridiculous!" shrieked the carpenter. "There aren't any trees in the desert!"
>
> "There aren't any, NOW," said Nasrudin calmly.

One of the questions then asks,

> "After Nasrudin finished work, he"

and provides as a distractor, "yelled at the carpenter." (The stipulated answer is told the carpenter what he had done.) This question is to test whether the child has absorbed the significance of the final three words, "said Nasrudin calmly," which appear after the punchline of the joke. In fact it seems that

these words were appended to the text expressly to set the trap for this question. But of course, competent readers use their understanding of the structure of genres to organize their memory of the contents of a text, and if there's one thing every competent reader or hearer knows about jokes, is that nothing is worth remembering after the punch line. Furthermore, it isn't obvious that this text is a joke until the word "NOW" is read, and so the IR will be doing a lot of heavy processing at this point. Moreover, the word in question is presented in capitals, which is a frequent orthographic device to indicate shouting. In sum, the passage + question seem explicitly designed to punish the test taker who allocates attention over the text in the way that a mature and competent reader would do.

Secondly, there is the very large subject of "test wisdom." It is quite clear that success on tests of reading comprehension depends to some considerable, if entirely unmeasured, degree on kinds of knowledge that have nothing to do either with ability to read or with knowledge of the world. A likely hypothesis--untested as far as we know--is that reading test scores could be raised substantially by teaching the children to read the questions before as well as after reading the text. This would of course be in violation of the test instructions and would also not indicate any increase in reading ability. One wonders to what uncontrolled extent some children now do this and what proportion of the variance in reading test scores actually recorded is accounted for by this variable alone. (Of course, if this should in fact be a better strategy, then there may be a confounding tendency for the children with high general intelligence to employ it.)

Related to the many issues of test wisdom is the problem of the "appropriate" use of background knowledge. We saw that in answering the "Bronco Buster" question about "keeping one hand free" the child is required to apply detailed schematic knowledge of the bronc riding contest and reason that the hand being held "free" must be held that way because that will make it harder to hold on--that is, the hand is "free" of anything attached to the horse that the rider might hold on to. On the other hand, we saw that in answering one of the questions about "Baseball" it is necessary for the test taker to suppress his or her knowledge of the fact that baseball is currently more popular than basketball in order to select a "correct" answer that says just the reverse.

The obvious retort to observations such as these is, "Yeah, but as competent adults we don't have any trouble figuring out when to use world knowledge and when not. So it's not an unreasonable standard to reward children for achieving." But although this retort is obvious, it is not to the point, because there is no assurance that an adult's ability to make the decisions that the testers reward is based on superior ability in reading. In fact it seems more likely that it is not so based, but rather based on longer experiences with this kind of test.

Test wisdom is not the basic object of the present study; rather we are more directly concerned with the construal principles used by good readers and the processes in which school age children acquire or fail to acquire these. Yet, although we cannot investigate the topic of test wisdom in depth (this is really the province of the educational psychologist), any careful analyst of the texts and questions put before children in tests of reading comprehension cannot fail to notice that these tests reward excessively test-taking strategies that are essentially unrelated to reading ability.

In subsequent sections, where we discuss the results of our interviews with third and fifth graders, we will have occasion to return to some of these considerations in a more pointed way.

APPENDIX A

Passages Discussed in the Report

"Amelia"

As she rolled to a stop in the center of the business district, she was quickly surrounded by townspeople. They were surprised when a woman stood up in the cockpit. When she raised her goggles they were even more surprised. She had been flying into the sun all day and her face was sunburned a bright red except for the white circles around her eyes where her goggles had protected her. She blinked out at the assembled population of Hobbs like a boiled owl.

- - - - - -

"Bananas"

Did you know that bananas don't grow on trees? Even though the banana plant looks like a tree, it has no trunks or limbs. Bananas are grown in South and Central America and India where the climate is hot and damp. They grow from flowers on the plant and are cut down before they get ripe. The banana plant dies down after just one crop. Since the banana does not have seeds like an apple or a pear, new plants are grown from the roots of the old plants.

- - - - - -

"Blue Egg"

Pat noticed it at the last possible moment. It was good that she had, because she had almost stepped on it. It was the most beautiful shade of blue she had ever seen. Pat decided that it must have fallen out of one of the nests in the tree above. She bent down to look at it and saw that it had not broken. "Oh good," she thought. "Maybe if I leave it alone, it will still hatch."

- - - - - -

"Bronco Buster"

If a bronco buster wants to win a rodeo contest, he must obey the contest rules. One of these rules is that the rider must keep one hand in the air. A rider who does not do this is disqualified.

- - - - - -

"Camping"

A strange noise wakes me up. I am very quiet, listening. I hear the wind blowing in the branches above the tent, and I can also hear footsteps. Some paper is rattled. All of a sudden a garbage can lid is tossed on the ground. I jump from my bed. My flashlight reflects two pairs of bears' eyes in the light beam, one from each trash can. This is so exciting to a city boy that I find it hard to go back to sleep in my sleeping bag.

- - - - - -

"Dear Dad"

Dear Dad,

We rode all day today. Mom parked the camper on a beach. We looked for shells and then fished. Mom caught a fish, and I cooked it for supper. We'll start out again in the morning. I can't wait to get to grandma's house tomorrow night!

Love,
Tam

- - - - - -

"Fritz"

Fritz lived in a neighborhood with many interesting people. But Fritz felt out of place. He felt that he was ordinary. In fact, he felt so ordinary that he thought nobody noticed him. Today I'll be different, he thought. I'll wear an extra hat.

Then he went for a walk. People smiled at him. But no one spoke. At last he asked a neighbor, "Don't you see something unusual about me today?"

"Yes," the neighbor said. "You're wearing three hats instead of your usual two."

- - - - - -

"Nasrudin"

The carpenter was astonished that such a weird, weak-looking creature as Nasrudin was applying for a job. "Okay, I'll give you a chance," said the doubtful carpenter finally. "Take this ax and chop as much lumber as you can." At dusk Nasrudin returned.

"How many trees have you felled?" questioned the carpenter.

"All the timber in the forest," Nasrudin replied.

Shocked, the carpenter glanced out his window. There were no trees standing on the hillside. Nasrudin had destroyed the entire forest. "Where did you learn to chop lumber?" asked the astonished carpenter.

"In the Sahara Desert," answered Nasrudin.

"That's ridiculous!" shrieked the carpenter. "There aren't any trees in the desert!"

"There aren't any, NOW," said Nasrudin calmly.

- - - - - -

"Shipwreck"

The boat was pounding on the rocks. The passengers were in an uproar. Women were screaming and men were yelling at the top of their voices. The captain of the boat went quietly from one place to another lowering the boats. If you looked at him you could never tell that the boat was about to sink.

- - - - - -

"Strange Machine"

In 1877, a machine appeared which surprised many people. Can you guess the name of this strange new machine? As you spoke into the mouthpiece and turned the handle, a tube covered with a thin piece of tin moved around. As the tube moved, a needle pressed deep lines into the tin. As you turned the handle once more, the needle touched against the same lines and played back your words. This was the first phonograph! How different from the hi-fi of today!

- - - - - -

References

de Beaugrande, Robert (1981) Design criteria for models of reading. Reading Research Quarterly 2. 261-313.

Brown, John Seely and R.B. Burton (1978) Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science 2. 155-192.

Fillmore, Charles and Paul Kay (1980) Progress Report: Text Semantic Analysis of Reading Comprehension Tests. (submitted to NIE in November, 1980)

Markman, E.M. (1979) Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child development 50. 643-655.

Spiro, Rand J.,Bertran Bruce and William F. Brewer (1980) Theoretical Issues in Reading Comprehension. Lawrence Erlbaum Associates, Hillsdale N.J.

Schank, Roger and Robert Abelson (1977) Scripts, Plans, and Understanding. Lawrence Erlbaum Associates, Hillsdale, N.J.

Wilensky, Robert (1982) Planning and Understanding. Lawrence Erlbaum Associates, Hillsdale, N.J.

Tests used as sources for items

CTB McGraw-Hill California Achievment Test Level 13:2.6-3.9; 15C: 4.6-5.9

CTBS (Comprehensive Test of Basic Skills) Level 1S: 2.5-4.9; 2S: 4.5-6.8

Gates-MacGinitie Primary C: 3.0-4.0

Metropolitan Achievment Tests, Form JI Elementary: 3.5-4.9; Intermediate:5.0-6.9

SRA Achievment Series Level D1: 3.5-4.5; Level E1: 4.5-6.5

Stanford Achievment Tests Primary Level II: 2.5-3.4; Primary Level III: 3.5-4.4; Intermediate Level I: 4.5-5.4; Level II: 5.5-6.9